SECOND EDITION

THE BACTERIOPHAGES

EDITED BY RICHARD CALENDAR

VISIT…

LANZAROTE
Caliente.COM

The Bacteriophages

This page intentionally left blank

The Bacteriophages

SECOND EDITION

Edited by
RICHARD CALENDAR

with editorial assistance by
Stephen T. Abedon

OXFORD
UNIVERSITY PRESS
2006

OXFORD
UNIVERSITY PRESS

Oxford University Press, Inc., publishes works that further
Oxford University's objective of excellence
in research, scholarship, and education.

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Copyright © 2006 by Oxford University Press, Inc.

Published by Oxford University Press, Inc.
198 Madison Avenue, New York, New York 10016

www.oup.com

Oxford is a registered trademark of Oxford University Press

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted, in any form or by any means,
electronic, mechanical, photocopying, recording, or otherwise,
without the prior permission of Oxford University Press.

Library of Congress Cataloging-in-Publication Data
The bacteriophages/[edited by] Richard Calendar.
p. cm.
Previously published: New York: Plenum Press, c1998.
Includes bibliographical references and index.
ISBN-13 978-0-19-514850-3
ISBN 0-19-514850-9
1. Bacteriophages. I. Calendar, Richard.
QR342.B335 2005
579.2'6–dc22 2004057572

9 8 7 6 5 4 3 2 1
Printed in the United States of America
on acid-free paper

Foreword

I know that I speak for both old and new devotees of bacteriophage when I say that this book has been wanted and needed for a very long time. An entire generation of graduate students has gone on to become department chairs since the last version of *The Bacteriophages*. In a real sense, the field of bacteriophage biology died, was buried and plowed under, but is now arising again as vigorous fresh green shoots from the soil so thoroughly enriched. The evidence of its death is indisputable. If you do a search of the NIH CRISP database for grants with "bacteriophage" in the title in the years of 1972 and 2002, the numbers come up approximately the same, above 200. However, a closer look reveals that most of the projects funded by NIH in 1972 involved research in which plaques were generated every week in the natural course of experimentation. When I looked through the grants funded in 2002, and put aside phage display and other library implementations, I could find fewer than 20 and I know for certain that several of these have expired since then. In the 1970s, the Cold Spring Harbor "Phage" meetings were basically all about some aspect of bacteriophage biology; now, the meeting is still affectionately called "Phage" but I can tell you as an organizer that it is a struggle to fill up even one evening session of a six-day conference with vaguely phage-related talks.

It should be a topic of some interest for science historians to explain how a field with so much momentum and so many talented practitioners suddenly turned its own lights off and just walked out the door. It was an exodus of talent and leadership of a scale, breadth, and suddenness never seen before in any field of biology, and perhaps in any field of modern science. My own theory is that the classical era of phage biology had at its core a suicidal impulse derived from physicists' reductionism. Others have suggested to me that the very success of molecular genetics, much of which was concerned with phages during the golden era, led to a kind of arrogance of invincibility and thus to a fearless rush to harvest the low hanging fruit in eukaryotic systems. Perhaps it also was alluring to be in the founding circle for new study sections, where presumably the competition would be much less intense.

In any case, there is a new phage biology emerging. In this new phage biology, the interest is in the phages themselves, not just in phages as a convenient system to learn new rules of molecular biology. The latter impulse is still alive, but its few remaining active adherents are mostly at or nearing retirement or bypass age. The new crowd of phage people come from all directions, not always intentionally. Phage are now being rediscovered as marvelous subjects for nanoscience (hardly a surprise to any kid who has ever seen a drawing of phage T4!). It also turns out, *mirabile dictu*, that phage are involved in many aspects of bacterial evolution and pathogenesis. Indeed, many diseases and most dissemination of virulence factors are basically phage phenomena, despite the decades-long aversion of funding agencies to consider phage as relevant to human disease. Moreover, phage are now being found to be sources of genetic information useful in combating drug-resistant pathogens, which should have been obvious long ago. In fact, much of this volume is written by members of the new wave. And, not least, phage are now being tamed and harnessed themselves as therapeutic agents, more than half a century after d'Herelle's lonely, ostracized demise.

Which brings us back to this book. Have you ever tried to find an up-to-date, comprehensive compendium of phage biology? Well, until now, you had very few choices, and most of them were out of print. As you will see, many of the chapters of this book are written by recognizable veterans of the classical phage years, but also many are written by new practitioners, some of whom didn't arrive at this juncture intentionally. They simply followed the science, and the science of microbiology is now coming back, full circle, to bacteriophage.

Ry Young

This page intentionally left blank

Preface

The idea of creating a book about the bacteriophages was conceived by the late Heinz Fraenkel-Conrat in his role as an editor of a series called The Viruses. The first edition of *The Bacteriophages* was published by Plenum Press in 1988 as two volumes. Its manager, Kirk Jensen, arranged for Oxford University Press to sponsor this second edition, and the new project was managed by Peter Prescott. Much has happened in the phage world since 1988. Horizontal gene transfer, which was just beginning to find acceptance then, has become an established issue today. Observation of the packaging of single DNA molecules and measurement of the force of the packaging motor is a recent happening. Phages with linear plasmid prophages were unknown in 1988. At that time, no one would have thought that a filamentous phage genome would integrate into a host genome. We have included articles on these subjects in this edition.

We aim to provide a current guide to each of the major phage families and to provide a general description of the kinds of phages that are associated with the major classes of eubacteria and archaea. In addition, we wish to highlight interesting current topics that are relevant to many of the phage families. I have been asked on several occasions to advise colleagues on the control of phage infections in bacterial cultures and fermentation facilities. To answer this need, Gregg Bogosian of Monsanto Corporation has provided a description of how phage control is achieved where its economic impact is high. Phages have been used widely to display antigens, and Björn Lindqvist has summarized the state of this field. Due to renewed interest in phage therapy, Carl Merril, Dean Scholl, and Sankar Adhya have contributed an article on the status and prospects of this art.

The latter stages of producing this volume were greatly aided by Steve Abedon (see phage.org, the google.com "phage ecology" *I'm feeling lucky* site). Steve came to *The Bacteriophages* first as an author (chapter 5, Phage Ecology) and then as developer of the companion web site (thebacteriophages.org). In the course of the latter he became intimately involved in the formatting and refinement of the figures (all of which may be found, some with color, at thebacteriophages.org). This activity led to editing of the figure legends, and then to editing of all 48 chapters. Closer to the beginning, Hans Ackermann suggested the order of presentation of articles.

This volume is dedicated to phage workers who have passed on recently: Gisela Mosig, who elucidated the diversity of DNA replication mechanisms used by coliphage T4; Edouard Kellenberger, whose expertise in electron microscopy led to the discovery of λ phage proheads, as well as to the superior electron microscope facilities at the European Molecular Biology Organization Laboratory; and Wolfram Zillig, who traveled the globe by airplane, foot, and cross-country skis to collect Archaea and show that they could release virus particles.

This page intentionally left blank

Contents

Foreword by Ry Young v
Contributors xi

PART I. General Background of Phage Biology

1. Phage and the Early Development of Molecular Biology 3
   William C. Summers
2. Classification of Bacteriophages 8
   Hans-W. Ackermann
3. Prophage Genomics 17
   Harald Brüssow
4. Evolution of Tailed Phages: Insights from Comparative Phage Genomics 26
   Harald Brüssow and Frank Desiere
5. Phage Ecology 37
   Stephen T. Abedon

PART II. Life of Phages

6. DNA Packaging in Double-Stranded DNA Phages 49
   Paul J. Jardine and Dwight L. Anderson
7. General Aspects of Lysogeny 66
   Allan Campbell
8. Gene Regulatory Circuitry of Phage λ 74
   John W. Little
9. Regulation of λ Gene Expression by Transcription Termination and Antitermination 83
   David I. Friedman and Donald L. Court
10. Phage Lysis 104
    Ry Young and Ing-Nang Wang

PART III. Cubic and Filamentous Phages

11. ϕX174 et al., the *Microviridae* 129
    Bentley A. Fane, Karie L. Brentlinger, April D. Burch, Min Chen, Susan Hafenstein, Erica Moore, Christopher R. Novak, and Asako Uchiyama
12. Filamentous Phage 146
    Marjorie Russel and Peter Model
13. PRD1: Dissecting the Genome, Structure, and Entry 161
    A. Marika Grahn, Sarah J. Butcher, Jaana K. H. Bamford, and Dennis H. Bamford
14. Lipid-Containing Bacteriophage PM2, the Type Organism of *Corticoviridae* 171
    Dennis H. Bamford and Jaana K. H. Bamford
15. Single-Stranded RNA Phages 175
    Jan van Duin and Nina Tsareva
16. Phages with Segmented Double-Stranded RNA Genomes 197
    Leonard Mindich

PART IV. Individual Tailed Phages

17. The T1-Like Bacteriophages 211
    Gregory J. German, Rajeev Misra, and Andrew M. Kropinski
18. T4 and Related Phages: Structure and Development 225
    Gisela Mosig and Fred Eiserling
19. Bacteriophage T5 268
    Jon R. Sayers
20. The T7 Group 277
    Ian J. Molineux
21. Bacteriophage N4 302
    Krystyna M. Kazmierczak and Lucia B. Rothman-Denes
22. Phage ϕ29 and its Relatives 315
    Margarita Salas
23. Bacteriophage SPP1 331
    Juan C. Alonso, Paulo Tavares, Rudi Lurz, and Thomas A. Trautner
24. Bacteriophage P1 350
    Hansjörg Lehnherr
25. The P2-Like Bacteriophages 365
    Anders S. Nilsson and Elisabeth Haggård Ljungquist
26. The Satellite Phage P4 391
    Gianni Dehò and Daniela Ghisotti
27. Bacteriophage λ and its Genetic Neighborhood 409
    Roger W. Hendrix and Sherwood Casjens
28. N15: The Linear Plasmid Prophage 448
    Nikolai V. Ravin

29. Bacteriophage P22 457
Peter E. Prevelige, Jr.
30. The Bacteriophage Mu 469
Luciano Paolozzi and Patrizia Ghelardini

PART V. Phages by Host or Habitat

31. Viruses of Archaea 499
Kenneth M. Stedman, David Prangishvili, and Wolfram Zillig
32. Phages of Cyanobacteria 517
Nicholas H. Mann
33. Marine Phages 534
Robert V. Miller
34. *Yersinia* Phages 545
Stefan Hertwig, Mikael Skurnik, and Bernd Appel
35. Temperate Bacteriophages of *Bacillus subtilis* 557
Pamela S. Fink and Stanley A. Zahler
36. Phages of *Lactococcus lactis* 572
Lone Brøndsted and Karin Hammer
37. The *Listeria* Bacteriophages 593
Martin J. Loessner and Richard Calendar
38. Mycobacteriophages 602
Graham F. Hatfull
39. Molecular Genetics of *Streptomyces* Phages 621
Margaret C. M. Smith
40. Mycoplasma Phages 636
Jack Maniloff and Kevin Dybvig
41. *Lactobacillus* Phages 653
Harald Brüssow and Juan E. Suárez

PART VI. Applications

42. Control of Bacteriophage in Commercial Microbiology and Fermentation Facilities 667
Gregg Bogosian
43. Phage-Based Expression Systems 674
Noreen E. Murray
44. Phage in Display 686
Bjorn H. Lindqvist
45. Bacteriophage as Pollution Indicators 695
Charles P. Gerba
46. The Use of Phage as Diagnostic Systems 702
Cath Rees
47. Bacteriophages in Bacterial Pathogenesis 710
Patrick L. Wagner and Matthew K. Waldor
48. Phage Therapy 725
Carl R. Merril, Dean Scholl, and Sankar Adhya

Index 743

Contributors

Stephen T. Abedon, Department of Microbiology, Ohio State University, Mansfield, OH 44906, USA

Hans-W. Ackermann, Department of Medical Biology, Faculty of Medicine, Laval University, Qc G1K7p4, Quebec, Canada

Sankar Adhya, Laboratory of Molecular Biology, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA

Juan C. Alonso, Departmento de Biotechnologia Microbiana, Centro Nacional de Biotecnologia-Consejo Superior de Investigaciones Cientificas, Campus Universidad Autonoma de Madrid, Cantoblanco, 28049 Madrid, Spain

Dwight L. Anderson, University of Minnesota, 18 256 Moos Tower, 515 Delaware Street S E, Minneapolis, MN 55455, USA

Bernd Appel, Department of Biological Safety, Horizontal Gene Transfer, Robert Koch Institut, Nordufer 20, 13353 Berlin, Germany

Dennis H. Bamford, Department of Biosciences and Institute of Biotechnology, Biocenter 2, FIN-00014, University of Helsinki, Finland

Jaana K. H. Bamford, Department of Biosciences and Institute of Biotechnology, Biocenter 2, FIN-00014, University of Helsinki, Finland

Gregg Bogosian, Monsanto Company, 800 N. Lindbergh Blvd, Creve Couer, MO 63167, USA

Karie L. Brentlinger, Department of Veterinary Science and Microbiology, University of Arizona, Building 90, Tucson, AZ 85721, USA

Lone Brøndsted, Department of Veterinary Pathobiology, Section for Microbial Food Safety, The Royal Veterinary and Agriculture University, Stigbøjlen 4, DK-1870 Frederiksberg C, Denmark

Harald Brüssow, Nestlé Research Center, 1000 Lausanne 26, Vers-chez-les-Blanc, Switzerland

April D. Burch, Department of Veterinary Science and Microbiology, University of Arizona, Building 90, Tucson, AZ 85721, USA

Sarah J. Butcher, Department of Biosciences and Institute of Biotechnology, Biocenter 2, FIN-00014, University of Helsinki, Finland

Richard Calendar, Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720-3202, USA

Allan Campbell, Department of Biological Sciences, Stanford University, Stanford, CA 94305, USA

Sherwood Casjens, Department of Pathology, University of Utah Medical Center, 30 North 1900 East, Salt Lake City, UT 84132, USA

Min Chen, Department of Veterinary Science and Microbiology, University of Arizona, Building 90, Tucson, AZ 85721, USA

Donald L. Court, Molecular Control and Genetics Section, Gene Regulation and Chromosome Biology Laboratory, Center for Cancer Research, National Cancer Institute, Building 539, P.O. Box B, Frederick, MD 21702-1201, USA

Gianni Dehò, Dipartimento di Scienze Biomolecolari e Biotecnologie, Università degli Studi di Milano, Via Celoria 26, 20133 Milan, Italy

Frank Desiere, Nestlé Reserch Center, 1000 Lausanne 26, Vers-chez-les-Blanc, Switzerland

Kevin Dybvig, Department of Genetics, University of Alabama at Birmingham, 720 South 20th Street, Kaul, Room 720, Birmingham, AL 35294, USA

Fred Eiserling, Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angles, Los Angeles, CA 90095, USA

Bentley A. Fane, Department of Veterinary Science and Microbiology, University of Arizona, Building 90, Tuscon, AZ 85721, USA

Pamela S. Fink, Bioprocess Development, Wyeth Research, Pearl River, NY 10965, USA

David I. Friedman, Department of Microbiology and Immunology, The University of Michigan, Medical School, Ann Arbor, MI 48109-0620, USA

Charles P. Gerba, Department of Soil, Water and Environmental Science, University of Arizona, Tucson, AZ 85721, USA

Gregory J. German, Faculty of Cellular and Molecular Biosciences, School of Life Sciences, Arizona State University, Tempe, AZ 85287-4501, USA

Patrizia Ghelardini, Istituto di Biologia e Patologia Molecolari del CNR, c/o Università di Roma "La Sapienza", Ple Aldo Moro 5, 00185 Rome, Italy

Daniela Ghisotti, Dipartimento di Scienze Biomolecolari e Biotecnologie, Università degli Studi di Milano, Via Celoria 26, 20133 Milan, Italy

A. Marika Grahn, Department of Biosciences and Institute of Biotechnology, Biocenter 2, 00014, University of Helsinki, Finland

Susan Hafenstein, Department of Veterinary Science and Microbiology, University of Arizona, Building 90, Tucson, AZ 85721, USA

Elisabeth Haggård-Ljungquist, Department of Genetics, Microbiology and Toxicology, University of Stockholm, S-106 91 Stockholm, Sweden

Karin Hammer, Microbial Physiology and Genetics, BioCentrum-DTU, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark

Graham F. Hatfull, Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260, USA

Roger W. Hendrix, Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260, USA

Stefan Hertwig, Department of Biological Safety, Horizontal Gene Transfer, Robert Koch Institut, Nordufer 20, 13353 Berlin, Germany

Paul J. Jardine, University of Minnesota, 18 256 Moos Tower, 515 Delaware Street S E, Minneapolis, MN 55455, USA

Krystina M. Kazmierczak, Department of Biology, Indiana University, Bloomington, IN 47405, USA

Andrew M. Kropinski, Department of Microbiology and Immunology, Queen's University, Kingston, Ontario K7L 3N6, Canada

Hansjörg Lehnherr, Department of Genetics and Biochemistry, Institute of Microbiology, Ernst Moritz Arndt University Greifswald, 17487 Greifswald, Germany

Björn H. Lindqvist, Department of Molecular Biosciences, University of Oslo, 0315 Oslo, Norway

John W. Little, Department of Biochemistry and Molecular Biophysics, University of Arizona, Tucson, AZ 85721, USA

Martin J. Loessner, Institute of Food Science and Nutrition, Swiss Federal Institute of Technology (ETH), Schmelzbergstrasse 7, LFV B20 CH-8092 Zurich, Switzerland

Rudi Lurz, Max-Planck-Institut für Molekulare Genetik, Ihnestrasse 73, D-1495 Berlin, Germany

Jack Maniloff, Department of Microbiology and Immunology, University of Rochester, New York, NY 14642, USA

Nicholas H. Mann, Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK

Carl R. Merril, Section on Biochemical Genetics, National Institute of Mental Health, National Institutes of Health, Bethesda, MD 20892, USA

Robert V. Miller, Department of Microbiology and Molecular Genetics, Oklahoma State University, Stillwater, OK 74078, USA

Leonard Mindich, Department of Microbiology, The Public Health Research Institute, Newark, NJ 07103, USA

Rajeev Misra, Faculty of Cellular and Molecular Biosciences, School of Life Sciences, Arizona State University, Tempe, AZ 85287-4501, USA

Peter Model, Laboratory of Genetics, The Rockefeller University, New York, NY 10021, USA

Ian J. Molineux, Department of Molecular Genetics and Microbiology, University of Texas, Austin, TX 78712-1095, USA

Erica Moore, Department of Veterinary Science and Microbiology, University of Arizona, Building 90, Tucson, AZ 85721, USA

Gisela Mosig, Formerly Department of Molecular Biology, Vanderbilt University, Nashville, TN 37235, USA

Noreen E. Murray, Institute of Cell and Molecular Biology, University of Edinburgh, Edinburgh EH9 3JR, UK

Anders S. Nilsson, Department of Genetics, Microbiology and Toxicology, University of Stockholm, S-106 91 Stockholm, Sweden

Christopher R. Novak, Department of Veterniary Science and Microbiology, University of Arizona, Building 90, Tucson, AZ 85721, USA

Luciano Paolozzi, Dipartimento di Biologia, Università di Roma "Tor Vergata," Via della Ricerca Scientifica, 00133 Rome, Italy

David Prangishvili, Institut Pasteur, 25–28 rue de Dr. Roux, 75724 Paris Cedex 15, France

Peter E. Prevelige, Jr. Department of Microbiology, University of Alabama at Birmingham, 845 19th Street South, BBRB 414, Birmingham, AL 35294-2170, USA

Nikolai V. Ravin, Center "Bioengineering", Russian Academy of Sciences, Prosp. 60-let Oktiabria, bld.7-1, Moscow 117312, Russia

Cath Rees, School of Biosciences, University of Nottingham, Sutton Bonington Campus, Loughborough, Leicestershire LE12 5RD, UK

Lucia B. Rothman-Denes, Department of Molecular Genetics and Cell Biology, University of Chicago, Chicago, IL 60637, USA

Marjorie Russel, Laboratory of Genetics, The Rockefeller University, New York, NY 10021, USA

Margarita Salas, Instituto de Biologia Molecular Eladio Vinuela (CSIC), Centro de Biologia Molecular Severo Ochoa (CSIC-UAM), Universidad Autonoma de Madrid, Cantobalanco, 28049 Madrid, Spain

Jon R. Sayers, Division of Genomic Medicine, Infection and Immunity, Medical School, University of Sheffield, Sheffield S10 2RX, UK

Dean Scholl, Section on Biochemical Genetics, National Institute of Mental Health, National Institutes of Health Bethesda, MD 20892, USA

Mikael Skurnik, Department of Bacteriology and Immunology, Haartman Institute, P.O. Box 63, 00014 Helsinki, Finland

Margaret C. M. Smith, University of Aberdeen, Institute of Medical Sciences, Aberdeen AB25 2ZD, UK

Kenneth M. Stedman, Biology Department, Portland State University, P.O. Box 751, Portland, OR 97207, USA

Juan E. Suárez, Area de Microbiologia, Universidad de Oviedo, Julian Claveria s/n, 33006 Oviedo, Spain

William C. Summers, Department of Therapeutic Radiology and Molecular Biophysics and Biochemistry, Yale University, 266 Whitney Avenue, New Haven, CT 06520-8114, USA

Paulo Tavares, Unité de Virologie Moléculaire et Structurale, UMR CNRS 2472-INRA 1157, Bâtiment 14B, CNRS, Avenue de la Terrasse, 91198 Gif-sur-Yvette Cédex, France

Thomas A. Trautner, Max-Planck Institut für Molekulare Genetik, Ihnestrasse 73, 14195 Berlin, Germany

Nina Tsareva, Depertment of Biochemistry, University of Leiden, Leiden, The Netherlands

Asako Uchiyama, Department of Veterinary Sciences and Microbiology, University of Arizona, Tucson, AZ 85721, USA

Jan van Duin, Department of Biochemistry, LIC, Leiden University, P.O. Box 9502, 2300 RA, Leiden, The Netherlands

Parteick L. Wagner, Department of Microbiology, Tufts University School of Medicine and Howard Hughes Medical Institute, 136 Harrison Avenue, Boston, MA 02111, USA

Matthew K. Waldor, Department of Microbiology, Tufts University School of Medicine and Howard Hughes Medical Institute, 136 Harrison Avenue, Boston MA 02111, USA

Ing-Nang Wang, Department of Biological Sciences, University at Albany, State University of New York, 1400 Washington Avenue, Albany, NY 12222, USA

Ryland F. Young III, Department of Biochemistry and Biophysics, Texas A&M University, 2128 TAMU, College Station, TX 77843-2128, USA

Stanley A. Zahler, Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA

Wolfram Zillig, Formerly Max-Planck-Institut für Biochemie, Am Klopferspitz 18A, 83152 Martinsried, Germany

This page intentionally left blank

PART I

GENERAL BACKGROUND OF PHAGE BIOLOGY

This page intentionally left blank

1

Phage and the Early Development of Molecular Biology

WILLIAM C. SUMMERS

Molecular biology evolved from multiple origins including the antivitalist biology of the early twentieth century which attempted to explain complex biological phenomena in terms of physical and chemical phenomena (16, 21), the Rockefeller Foundation program, led by physicists and mathematicians who believed the "human sciences" were ripe for deeper understanding based on chemistry and physics (29), research by physicists and chemists who saw life as a challenge to their burgeoning understanding of the fundamental structure of matter (2, c22), and work by a few microbiologists and geneticists who sought better understanding of the natures of genes and microbes (23).

Although on one hand profoundly reductionistic, molecular biology eventually drew its boundaries to encompass organizational principles of cells and even organisms. These research pathways took two distinct routes starting in the 1930s. One route employed the new macromolecular chemistry of proteins pioneered by Svedberg, Pauling and like-minded chemists, and X-ray diffraction analyses being developed by the mineralogists and physicists such as the Braggs, father and son, and Astbury and colleagues, mainly in Great Britain. Another route was taken by biologists and some curious physicists who approached complex biological problems with a combination of reductionism and holism.

The latter program was reductionistic because it endeavored to find "simple systems" as exemplars of all of biology, but holistic because it treated the entire biological system (e.g., a bacterial cell or virus) as a "black box" with its intrinsic complexities. An article of faith in this endeavor, however, was that the observable behavior of the black box reflected some of the simple, knowable processes that occurred inside. The goal of this research program was to design experiments that connected empirical observations to fundamental cellular processes.

Heredity and "the Gene"

A crucial biological problem recognized by some scientists in the 1930s involved the mechanisms of heredity. Genetic analysis of model organisms such as insects, especially fruit flies, and plants such as barley and maize suggested that "the gene" as a concept was far from clear. On one hand, genes were transmitted from generation to generation with remarkable fidelity, and on the other hand, genes were implicated in the development of organisms and cells. The unusual stability of genes, however, contrasted with their rare but dramatic mutation to other forms which were (usually) highly stable. These linked puzzles of faithful transmission and rare but stable mutation suggested that genes were objects with unusual properties. Prior to the current understanding of genes as macromolecular entities with information-containing structure, genes were conceived as "factors," "determinants," or simply conceptual entities without physical reality at all (9, 11).

While genes were believed to be associated with the visible cytological structures called chromosomes, the physical nature of genes was unclear. The concept of the gene as some sort of "unit of hereditary information" was not in the thinking of biologists prior to the late 1950s (12). In the 1920s and 1930s, the diversity of proteins, deduced from the amino acid compositions and sizes of proteins from different sources, suggested that proteins might be related to the remarkable diversity and specificity attributed to genes.

Bacteriophages and Genes

Almost from their first discovery in 1917, bacteriophages were viewed as relevant to research on these studies on the nature of the gene. Because of their small size (and presumed simple structure) and the claim that they multiplied to yield progeny similar to the parental phage, bacteriophages were viewed by some investigators as the most primitive form of life and, perhaps, "naked genes" (20). Although it seems likely that bacteriologists had encountered bacteriophages from the beginning of their discipline, it was in 1915 that F. W. Twort reported on a strange phenomenon he termed "glassy transformation" of cultures of micrococci (27). He noted that the calf lymph, from which he was trying to grow vaccinia on cell-free medium, contained a serially transmissible agent that induced a watery dissolution of the bacteria, leaving nothing but subcellular granules. Quite independently, and for quite different reasons, Felix d'Hérelle in 1917 discovered a microbe that was "antagonistic" to bacteria and that resulted in their lysis in liquid culture and killing in discrete patches (he called them plaques) on the surface of agar seeded with the bacteria. d'Hérelle conceived of these invisible microbes as "ultraviruses" that invaded bacteria and multiplied at their expense, and so he termed them bacteriophage (7). He astutely realized that the plaque count provided a way to enumerate these invisible agents, which he conceived as particulate. He was able to show that phage multiplied in "waves" or "steps" representing cycles of infection, multiplication, release, and reinfection.

d'Hérelle pioneered two important pathways of phage research: one was based on his finding that phage titers rose in patients with infectious diseases just as recovery was taking place. From these observations he believed that phages represented a natural agent of resistance to infectious diseases, and went on to advocate phages as therapeutic agents in the pre-antibiotic era. His concept of phages as viruses of bacteria was not widely accepted (the leading authorities up until the early 1940s thought that the lytic phenomena associated with bacteriophages resulted from an autocatalytic activation of an induced endogenous lytic enzyme). To counter his critics as well as to establish his priority for the discovery of phage in a long-running dispute with Twort's supporters, d'Hérelle's second research program examined the biological nature of bacteriophage. All his evidence pointed to the conception of phages as organized infectious agents that are obligate intracellular parasites. The antigenic properties as well as host ranges of phages appeared to be characteristic of given "races" of phages, and gave hints that phage genetics might be a fruitful topic for study (8).

Genes, Phage, and Radiation

For most of the first half of the twentieth century the chemical structures and functions of most cellular components were inaccessible to direct experimentation. During this period, however, quite ingenious use was made of a technique that was appropriated directly from the field of atomic physics. Just as the bombardment of atomic nuclei by high-energy particles such as X-rays and electrons provided information about the internal structure of the atom, by the mid-1920s physicists had developed conceptual approaches to probe the interior of cells by bombardment with projectiles such as X-rays and other high-energy particles. From a simple X-ray inactivation curve, without any chemical or structural knowledge, the size and shape of an enzyme or virus could be deduced. When it was shown by Herman J. Muller in 1922 that X-rays could induce mutations in genes, the target theory approach was immediately applied to genes to determine the basic size and shape of "a gene." Because this approach was well known to physicists and employed straightforward mathematical formalisms, it is not surprising that many physicists who were interested in biological problems used radiobiological methods in their research.

For example, one of the century's greatest physicists, who thought seriously about biological problems, was Niels Bohr. His essay, "Light and Life," published in 1933 (2), was said to have been highly influential on the thinking of many physicists. Leo Szilard, the inventor of the nuclear chain-reaction, became involved in full-time radiobiological research after World War 2 when he decided to take up biological problems after his wartime work on nuclear physics. In France, Marie Curie undertook biological studies of radiation based on target theory models, and Fernand Holweck, another brilliant physicist of broad interests, collaborated with biologists at the Institut Pasteur, including Salvador Luria, a medically trained physicist working first in Rome, then in Paris, and finally in the United States, to study virus structure and function in the late 1930s using formal target theory.

The application of radiobiology to study the nature of the gene was the main program of a small group in pre-war Germany consisting of Max Delbrück, a physicist, Karl Zimmer, a radiobiologist, and Nicolai Timoféeff-Ressovsky, a visiting Russian geneticist (26). These scientists published what Gunther Stent has characterized as nothing less than an attempt at a quantum mechanical description of the gene (25). Their experimental approach was, again, almost exclusively based on target theory.

What started as a problem in microbiology was adopted and adapted to the research program of the radiobiologists and geneticists (24). Salvador Luria studied the sizes of phages using target theory models. Delbrück, a student of Meitner and Hahn, moved to the California Institute of Technology in 1938 with plans to develop his theories of the gene using fruit flies, but upon arriving at Caltech he learned of phage research already being carried out there by Emory Ellis. Ellis had initiated a project to study the basic biology of phage as a model for oncogenic virus research. Delbrück was impressed by Ellis's confirmation of d'Hérelle's one-step growth curves and so he joined in the study of bacteriophage multiplication as a black box model (or "gadget" in Delbrück's terminology) for heredity.

Concurrently with the work of Ellis and Delbrück and that of Luria, Alfred Hershey was studying phages from the point of view of their physiology. He was collaborating with Jacques Bronfenbrenner at Washington University in St. Louis, Missouri who had a long interest in the possible metabolic and structural organization of bacteriophage.

Delbrück was a natural organizer and, together with Luria and Hershey, he began to recruit people to work on phage biology. He developed a group of protégés, followers, devotees, and students who adopted his viewpoints on the important problems of phage research and the legitimate ways to approach these problems. Delbrück saw the value in focusing research on a small group of phages so that results from different laboratories could be compared, and so he selected a group of "authorized phages" which became designated the T-phages, T1–T7 (T for type) (6).

Gene Regulation and Lysogenic Phages

While Delbrück was suspicious of the phenomenon known as lysogeny, this problem was directly attacked by André Lwoff and his colleagues in Paris in the 1950s. Lysogenic phenomena were observed almost from the first days of phage research, but the nature of the relationship between phage and host was unclear. Was lysogeny a sort of smoldering, persistent infection with the phage muliplying in some sort steady state with the growth of the host, or did the phage become truly latent? Eugène and Elizabeth Wollman suggested that phage in the lysogenic state behaved as part of the cellular hereditary apparatus. Lwoff and Antoinette Gutmann finally clarified the nature of lysogeny and christened the latent form of the phage "prophage" in their work in 1950, in which they followed phage induction and release from single cells using direct microscopic observation and sampling with a micromanipulator (18). Lysogeny, induction, and its regulation became a major focus of phage research in Paris in the 1950s. The thesis of François Jacob was on lysogeny in *Pseudomonas pyocyanea*, and in the 1950s the study of lysogeny provided the groundwork for the operon concept of gene regulation that was developed in the Service de Physiologie Microbienne at the Institut Pasteur (19).

Just as the T-phages were the model organisms for lytic phage research, bacteriophage λ, discovered in 1951 by Esther Lederberg, became the prototypic lysogenic phage (13). Study of λ phage has provided a deep understanding of regulation of gene expression on one hand, and the mechanisms of lysogeny on the other.

While the French phage workers pursued research of a more physiological sort, based as it was on Lwoff's life long interest in growth, nutrition, and physiological adaptations, Delbrück's followers, who came to be called the American Phage Group, favored more direct physical approaches, one of which was the target theory method.

Phage and the Physical Gene

The nature and cause of gene mutations were investigated in the 1930s, and a key question emerged from this work: Were mutations caused by the selective growth conditions (e.g., the addition of lactose to the medium) or did the mutations occur randomly all the time and their existence become known by imposition of the selective growth conditions? The outcome of this research would have profound implications not only for the science of genetics but also for a deeper understanding of evolutionary biology in general.

The major problem in this work was one of experimental design: how to observe a rare event that happened in a huge population prior to the selection for the outcome of that event. In a particularly clear and convincing work, Isaac M. Lewis examined the mutational change from the inability to ferment lactose to the ability to use this sugar source in the *Escherischia coli* strain *mutabile*. He concluded that the mutations to lactose utilization occurred prior to the selection, not as a consequence of exposure to the selective conditions. His elegant and clear approach, however, did not change many minds, but in the 1940s two related experimental approaches gave results believed to settle this question. Both of these employed bacteriophage as experimental tools. Delbrück and Luria knew from the work of d'Herelle that bacteria often developed heritable resistance to phage. Their routine use of statistical models in their target theory, as well as their backgrounds in atomic physics, helped them to devise a statistical approach ("fluctuation test") to show that phage-resistant mutants existed in the bacterial population prior to exposure to the lethal effects of phage (17). The method of Luria and Delbrück, and a related procedure devised in 1949 by Howard B. Newcome, were indirect and mathematical. However, in 1952 Joshua and Esther Lederberg devised a simple and direct way to demonstrate that mutations were occurring in a random way, independent of the selection procedures: they transferred very large numbers of colonies from one plate to another by the use of velvet cloth as a transfer tool (15). Thus, "replica plates" could be used to test colonies in great numbers for mutant properties. They applied this technique to study of phage resistance as well as streptomycin resistance: again, it was clear that the mutants had appeared before the application of the selective agent.

The random nature of mutation and its very low frequency of occurrence suggested that it might be similar to, or governed by, a quantized, two-state process. This model appealed to physicists such as Delbrück, Erwin Schroedinger and Bohr who thought deeper understanding of this paradoxical behavior might reveal new physical laws of nature (25). Remarkably, this deeper insight was provided by strictly formal genetic analysis of phage mutations. Hershey had employed plaque morphology mutants (large plaques, interpreted as rapid-lysis mutants or "r-mutants") to show that phages could be "mated" or "crossed" by simultaneous mixed infections and thus it was possible to carry out formal genetic analysis on phage just as with sexually mating organisms (10). Seymour Benzer found one class of Hershey's r-mutants of phages T2 and T4 to be particularly interesting: they did not give any plaques at all on bacterial hosts carrying the λ prophage (K-12 λ) but gave the large (r-mutant) plaques on the usual bacterial host (strain B). Benzer exploited this case of conditional expression of a phage mutation to develop the fine structure genetic analysis of the *T4rII* gene (1). Because of the strong specificity of the system (the discrimination against the rII mutants in host K-12 λ is greater than 10^8), very low frequency wild-type r^+ recombinants or revertants can be detected, and hence even recombination between very close mutations can be observed. Benzer calculated that he could detect recombination between adjacent base pairs in the rII gene. The detailed genetic analysis of deletions and insertion mutations in the *T4rII* gene also provided strong evidence for the triplet nature of the genetic code in phage crosses by Francis H. C. Crick, Leslie Barnett, Sydney Brenner, and Richard J. Watts-Tobin (5).

By the 1960s it became clear that phages represented only one sort of extrachromosomal genetic structure that could exist in bacteria. The fertility factor F, transmissible drug-resistant determinants, as well as prophages all represented examples of what Joshua Lederberg termed "plasmids" (14). It remained to be determined in what form these plasmids existed in the cell. Some experiments suggested that the plasmids (the F-factor in Hfr strains, and some prophages) were attached to the cell chromosome somehow. A very fruitful model for how this attachment could occur was proposed in 1962 by Allan Campbell (3). He reasoned that because the genetic map of phage T4 was circular while the phage DNA appeared to be linear, the phage might assume a circular intracellular form. This circular form could, with a single reciprocal recombination event, become linearly integrated into the chromosomal DNA. Although it turned out that T4 has a circular genetic map for reasons other than forming a physically circular model, Campbell's model was soon confirmed for phage λ, and eventually many other plasmids. It has provided the conceptual basis for retrovirus integration and excision in animal cells as well.

Genes as Information

As the study of phage replication and gene expression developed, it became clear that the full understanding of the workings of genes could be seen as a unified process, which was captured by Crick's metaphor of the so-called central dogma of molecular biology. Describing the function of genes in information-theoretic terms, he stated that information flows from DNA to DNA and from DNA to RNA and thence to proteins. He also noted at the time that it was theoretically possible for information flow to occur from RNA to DNA and RNA to RNA (4). Key experimental support for the role of RNA as an intermediary in such information transfer was the finding that in phage-infected cells, the base composition of newly synthesized RNA was much more like that of the phage DNA than the host DNA (28). Thus, it was concluded that phage gene function required the synthesis of RNA molecules that differed from those present in the uninfected cell, and that the RNA base composition, and hence its information content, was determined by the DNA of the phage.

From the 1960s, phage research was greatly advanced by new biochemical approaches originating from the early work of Seymour Cohen, Lloyd Kozloff, and others. The availability of radioactive isotopic techniques, improved methods of protein chemistry, and advances in enzymology contributed to the influx of biochemists and biochemical methods in phage work. This fruitful collaboration between biochemists, geneticists, and microbiologists, led to detailed descriptions of the mechanisms of phage replication and transcription, and of phage morphogenesis and assembly, and the detailed understanding of phage adsorption and entry phenomena. Many of these advances are still ongoing and are part of the recent history of phage molecular biology described in this book.

References

1. Benzer, S. 1955. Fine structure of a genetic region in bacteriophage. Proc. Natl. Acad. Sci. USA. 41:344–354.
2. Bohr, N. 1933. Light and life. Nature 131:421–423, 457–459.
3. Campbell, A. M. 1962. Episomes. Adv. Genet. 11:101–145.
4. Crick, F. H. C. 1970. Central dogma in molecular biology. Nature 227:561–563.
5. Crick, F. H. C., L. Barnett, S. Brenner, and R. J. Watts-Tobin. 1961. General nature of the genetic code for proteins. Nature 192:1227–1232.
6. Demerec, M., and U. Fano. 1945. Bacteriophage-resistant mutants in *Escherichia coli*. Genetics 30:119–136.
7. d'Hérelle, F. 1917. Sur un microbe invisible antagoniste des bacilles dysentériques. C. R. Acad. Sci. 165:373–375.
8. d'Herelle, F. 1931. Bacterial mutations. Yale J. Biol. Med. 4:455–461.
9. Falk, R. 2000. The gene — a concept in tension, p. 317–348. *In* P. J. Beurton, R. Falk, and H.-J. Rheinberger (eds.) The Concept of the Gene in Development and Evolution: Historical and Epistemological Perspectives. Cambridge University Press, Cambridge, UK.
10. Hershey, A.D., and R. Rotman. 1949. Genetic recombination between host-range and plaque-type mutants of bacteriophage in single bacterial cells. Genetics 34:44–71.
11. Kitcher, P. 1982. Genes. Br. J. Phil. Sci. 33:337–359.
12. Kay, L. 1999. Who Wrote the Book of Life? Stanford University Press, Stanford, Calif.
13. Lederberg, E. M., and J. Lederberg. 1953. Genetic studies of lysogenicity in *Escherichia coli*. Genetics 38:51–64.

14. Lederberg, J. 1952. Cell genetics and hereditary symbiosis. Physiol. Rev. 32:403–430.
15. Lederberg, J., and E. M. Lederberg. 1952. Replica plating and indirect selection of bacterial mutants. J. Bacteriol. 63:399–406.
16. Loeb, J. 1912. The Mechanistic Conception of Life. University of Chicago Press, Chicago, Ill.
17. Luria, S. E., and M. Delbrück. 1943. Mutations of bacteria from virus sensitivity to virus resistance. Genetics 28:491–511.
18. Lwoff, A., and A. Gutmann. 1950. Recherches sur un Bacillus Megatherium lysogène. Ann. Inst. Pasteur 78: 711–739.
19. Morange, M. 2002. Les mousquetaires de la nouvelle biologie: Monod, Jacob, Lwoff. Pour la Science 10:3–96.
20. Muller, H. J. 1922. Variation due to change in the individual gene. Am. Nat. 56:48–49.
21. Needham, J. 1931. Chemical Embryology. Cambridge University Press, Cambridge, UK.
22. Schrödinger, E. 1945. What Is Life? Cambridge University Press, Cambridge, UK.
23. Summers, W. C. 1991. From culture as organism to organism as cell: historical origins of bacterial genetics. J. Hist. Biol. 24:171–190.
24. Summers, W. C. 1993. How bacteriophage came to be used by the Phage Group. J. Hist. Biol. 26:255–267.
25. Stent, G. S. 1992. Introduction: Waiting for the paradox, p. 3–8. *In* J. Cairns, G. S. Stent, and J. D. Watson (eds.), Phage and the Origins of Molecular Biology (expanded edition). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
26. Timoféeff-Ressovsky, N. W., K. G. Zimmer, and M. Delbrück. 1935. Über die Natur der Genmutation und der Genstruktur. Nachrichten Gesellschaft der Wissenschaften zu Göttingen, math-phys Kl. Fachgruppe 6:189–245.
27. Twort, F. W. 1915. An investigation on the nature of the ultra-microscopic viruses. Lancet II:1241–1243.
28. Volkin, E., and L. Astrachan. 1956. Phosphorus incorporation in *Escherichia coli* ribonucleic acid after infection with bacteriophage T2. Virology 2:149–161.
29. Weaver, W. 1970. Scene of Change: A Life in American Science. Scribner, New York.

2

Classification of Bacteriophages

HANS-W. ACKERMANN

Bacteriophages or "phages" were discovered twice in a short time. The British pathologist Frederick William Twort described in 1915 the glassy transformation of "*Micrococcus*" colonies by an unknown agent. Félix Hubert d'Hérelle, a French Canadian then working at the Pasteur Institute of Paris, observed the destruction of *Shigella* bacteria in broth (5). Contrary to Twort, he clearly recognized the viral nature of this phenomenon and devoted the rest of his scientific career to it. He coined the term "bacteriophage," devised several techniques still in use, and introduced the phage treatment of infectious diseases. For him, there was only one bacteriophage species with many races: the *Bacteriophagum intestinale* (6).

The viral nature of bacteriophages was recognized in 1940 with the advent of the electron microscope. In 1962, Lwoff, Horne, and Tournier published a system of viruses based on morphology and nucleic acid type. They proposed the order *Urovirales* for tailed phages and the families *Inoviridae* and *Microviridae* for filamentous and ϕX-type phages, respectively (9). A further milestone was the recognition of six basic phage types: tailed phages, filamentous phages, and icosahedral phages with single-stranded DNA or single-stranded RNA. This simple scheme, proposed by Bradley in 1967 (4), is still the basis of the present edifice of phage classification.

In 1971, the International Committee on Taxonomy of Viruses (ICTV) classified phages into six genera corresponding to five of Bradley's basic types, namely the T4, λ, ϕX174, MS2, and fd phage groups, and the newly described type PM2 (16). New phage groups were added over time. The most recent development is the establishment of the order *Caudavirales* for tailed phages and of 15 tailed phage genera (1, 10, 14). At present, at least 5136 bacterial viruses have been examined in the electron microscope (3). This makes bacteriophages the largest viral group in nature.

Phage Classification Today

The ICTV presently recognizes one order, 13 families, and 31 genera of phages (14). Virions have binary, cubic, or helical symmetry, or are pleomorphic. Most phages contain dsDNA, but there are small phage groups with single-stranded (ss) DNA, ssRNA, or double-stranded (ds) RNA. A few types have a lipid-containing envelope or contain lipids as part of the particle wall. Tailed phages (binary symmetry) total about 5100 viruses at the time of writing (96% of phages) and are classified into the order *Caudovirales* and three very large, phylogenetically related families. By contrast, "cubic," filamentous, and pleomorphic phages comprise less than 190 viruses only (3.6% of phages) and are classified into 10 small families. They are extremely diversified by their basic properties and seem to constitute many different lines of descent. Capsids with cubic symmetry are icosahedra or related bodies. Particles are enveloped or not. The presence of lipids is accompanied by low buoyant density and high sensitivity to chloroform and ether.

As elsewhere in virology, families are chiefly defined by nature of nucleic acid and particle morphology (figure 2-1; table 2-1). There are no universal criteria for genus and species delineation. The ICTV uses every available property for classification and has adopted the "polythetic species concept," meaning that a species is defined by a set of properties, some of which may be absent in a given member (13). Taxonomic names of orders, families, and genera are typically constructed from Latin or Greek roots and end in *-virales*, *-viridae*, and *-virus*, respectively. Most taxa of "cubic," filamentous, and pleomorphic phages have latinized names. So far, tailed phage genera have vernacular names only.

Viruses with Binary Symmetry (Caudovirales, Tailed Phages)

At least 4950 tailed phages are known (3). Particles consist of a head with cubic symmetry and a "helical" tail and are said to be of binary symmetry (9). The head–tail structure is unique in virology. Although tail-like structures occur in a few other viruses, such as tectiviruses (see below), some algal viruses, and polydnaviruses, they are inconstant and

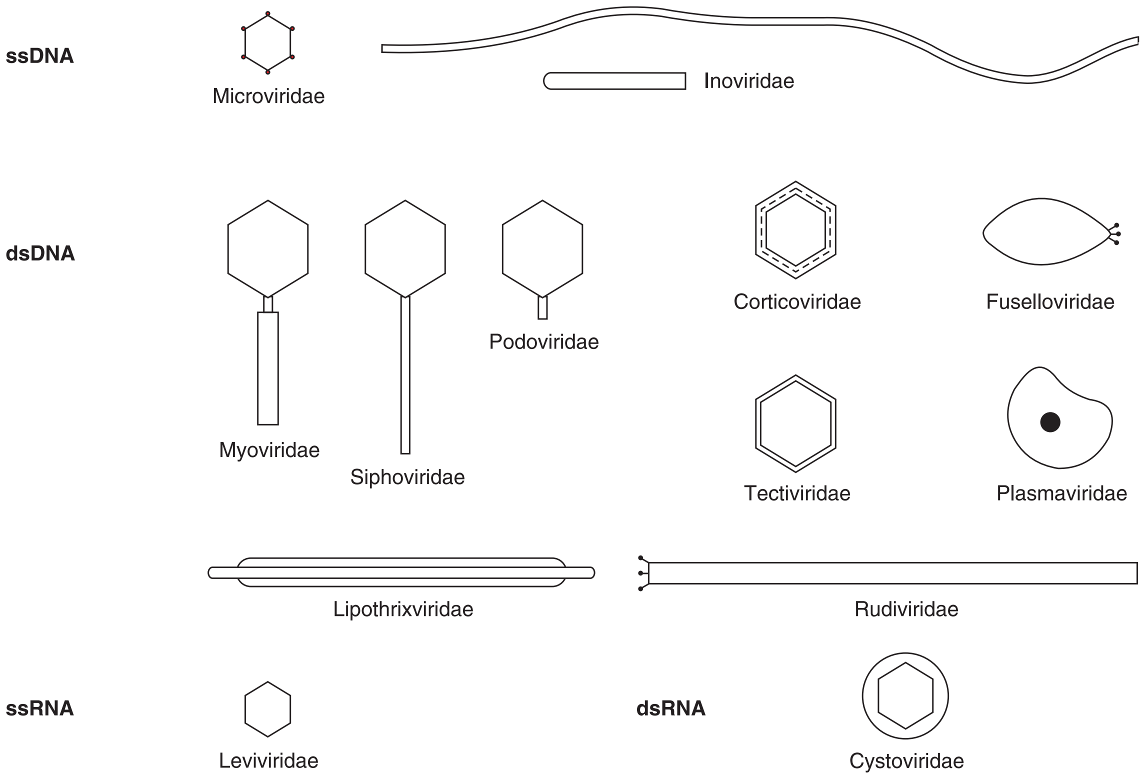


Figure 2-1 Schematic representation of major phage groups.

Table 2-1 Classification and Basic Properties of Bacteriophages

Symmetry	Nucleic acid	Order and families	Genera	Members	Particulars
Binary (tailed)	DNA, ds, L	*Caudovirales*	15	4950	
		Myoviridae	6	1243	Tail contractile
		Siphoviridae	6	3011	Tail long, noncontractile
		Podoviridae	3	696	Tail short
Cubic	DNA, ss, C	*Microviridae*	4	40	
	ds, C, T	*Corticoviridae*	1	3?	Complex capsid, lipids
	ds, L	*Tectiviridae*	1	18	Internal lipoprotein vesicle
	RNA, ss, L	*Leviviridae*	2	39	
	ds, L, S	*Cystoviridae*	1	1	Envelope, lipids
Helical	DNA, ss, C	*Inoviridae*	2	57	Filaments or rods
	ds, L	*Lipothrixviridae*	1	6?	Envelope, lipids
	ds, L	*Rudiviridae*	1	2	Resembles TMV
Pleomorphic	DNA, ds, C, T	*Plasmaviridae*	1	6	Envelope, lipids, no capsid
	ds, C, T	*Fuselloviridae*	1	8?	Spindle-shaped, no capsid

Modified from Ackermann, H.-W. 2001. Le matin des bactériophages. Virologie 5:35–43. With permission of John Libbey-Eurotext. Phage numbers are from (3).

C, circular; L, linear; S, segmented; T, superhelical; ss, single-stranded; ds, double-stranded.

do not compare with the regular and constant tails of tailed phages (1). This feature, and many morphological or physiological properties, indicate that tailed phages constitute a monophyletic evolutionary group. At the same time, tailed phages are extremely varied in DNA content and composition, dimensions and fine structure, physiology, and lifestyle; for example, DNA sizes vary between 17 and 500 kb and tail lengths range from 10 to 800 nm. Tailed phages represent the most diversified of all virus groups.

Virions have no envelope and consist typically of protein and DNA. Lipids are generally absent. Heads are icosahedra or elongated derivatives thereof. Isometric heads prevail (85%). Capsomers are rarely visible. Tails are helical or consist of stacked disks and carry in most cases fixation structures such as base plates, spikes, or terminal fibers. Heads, tails, and tail fibers are synthesized in separate pathways and assembled later. Virion response to inactivating agents is variable and no generalization is possible here.

Table 2-2 Selected Properties of Major Phage Groups

Family or group	Examples	Capsid (nm)	Lipids (%)	Nucleic acid		Infection	Release
				%	MW		
Caudovirales	T4, λ, T7	67	–	46	79	V or T	Lysis
Range		30–160		30–62	17–498		
Microviridae	ϕX174	27	–	26	4.4–6.1	V	Lysis
Corticoviridae	PM2	60	13	14	9.0	V	Lysis
Tectiviridae	PRD1	63	15	14	15	V	Lysis
Leviviridae	MS2	23	–	30	3.5–4.3	V	Lysis
Cystoviridae	ϕ6	75–80	20	10	13.4	V	Lysis
Inoviridae, Inovirus	fd	760–1950 × 7	–	6?–21	5.8–7.3	S or T	Excretion
Plectrovirus	L51	85–250 × 7	–		4.4–8.3	P	Excretion
Lipothrixviridae	TTV1	400–2400 × 20–40	22	3	16–42	V or T	Lysis
Rudiviridae	SIRV1	780–900 × 23	–		33–36	P	Excretion
Plasmaviridae	L2	80	11		11.7	T	Excretion
Fuselloviridae	SSV1	85 × 55	10		15	T	Excretion

Modified from Ackermann, H.-W. 2001. Le matin des bactériophages. Virologie. 5:35–43. With permission of John Libbey-Eurotext. More physiological data may be found in (2).

MW, molecular weight in kb or kpb × 10^6; P, permanent; S, steady-state, T, temperate; V, virulent; –, none.

Despite the general absence of lipids, about one third of tailed phages are chloroform-sensitive.

The DNA is a single, linear, double-stranded filament. Its composition generally reflects that of the host bacterium, but some DNAs contain unusal bases such as 5-hydroxymethylcytosine or 5-hydroxymethyluracil. Genetic maps are complex and include about 290 genes in phage T4 (possibly much more in larger phages). Genes for related functions cluster together. During maturation, the DNA enters preformed capsids.

Tailed phage are divided into three families:

1. *Myoviridae*, with contractile tails consisting of a sheath and a central tube (25% of tailed phages). The sheath is separated from the head by a neck.
2. *Siphoviridae*, with long, noncontractile tails (61%).
3. *Podoviridae*, with short tails (14%).

Genera are differentiated by genome structure (presence or absence of *cos* or *pac* sites, terminal redundancies, and circular permutations), concatemer formation, presence or absence of unusual bases and DNA or RNA polymerase genes, and DNA sequence. Classification into genera is still at its beginnings (table 2-3). Many more tailed phage genera are likely to be recognized in the future. The few presently defined genera may be seen as islands in a sea of nonclassified phages, or as crystallization points for phages awaiting classification. The term "λ-like viruses" is not a synonym of "lambdoid phages." The latter is a vernacular term that denotes the presence of common genes. It may be recalled that λ and P22, here attributed to distinct genera, have no more than 13.5% DNA homology (12, 15). About 250 species are presently recognizable, mostly on the basis of morphology, DNA–DNA hybridization and sequencing, and serology.

Phages with Cubic Symmetry and DNA

Microviridae (ssDNA)

Virions are small (27 nm in diameter), have no envelope, and contain a single piece of circular ssDNA. Phages infect very different hosts (enterobacteria, *Bdellovibrio*, *Chlamydia*, *Spiroplasma*) and are classified into four genera.

Corticoviridae (dsDNA)

The only certain member of the family *Corticoviridae* is a maritime phage, PM2. Its capsid consists of two protein shells between which a lipid bilayer is sandwiched. Two similar viruses were isolated from seawater, but their taxonomic position is unclear.

Tectiviridae (dsDNA)

Phages are unique by their structure and mode of infection. A rigid outer protein capsid surrounds a thick, flexible lipoprotein vesicle. Upon adsorption to bacteria or chloroform treatment, this vesicle becomes a taillike tube about 60 nm in length, thus a nucleic acid ejection device. Tectiviruses of bacilli have apical spikes. Despite their small numbers, tectiviruses occur in widely different hosts such as enterobacteria, bacilli, and *Thermus* bacteria.

Phages with Cubic Symmetry and RNA

Leviviridae (ssRNA)

Virions of the *Leviviridae* resemble enteroviruses and have no particular morphological characteristics. Most

Table 2-3 Tailed Phage Genera

Family	Genus	Type species	Species	Members[a]	Principal hosts
Myoviridae	T4-like viruses	T4	7	47(+100)	Enterobacteria
	P1-like viruses	P1	3	12	Enterobacteria
	P2-like viruses	P2	2	16	Enterobacteria
	Mu-like virues	Mu	1	2	Enterobacteria
	SPO1-like viruses	SPO1	1	13	*Bacillus*
	ΦH-like viruses	ΦH	1	2	*Halobacterium*[b]
Siphoviridae	λ-like viruses	λ	1	7	Enterobacteria
	T1 -like viruses	T1	1	11(+50)	Enterobacteria
	T5-like viruses	T5	1	5(+20)	Enterobacteria
	L5-like viruses	L5	1	4(+15)	*Mycobacterium*
	c2-like viruses	c2	1	5(+200)	*Lactococcus*
	ΨM1-like viruses	ΨM1	1	3	*Methanobacterium*[b]
Podoviridae	T7-like viruses	T7	3	26	Enterobacteria
	P22-like viruses	P22	1	11	Enterobacteria
	φ29-like viruses	φ29	4	12	*Bacillus*

[a] Parentheses indicate approximate numbers of poorly characterized isolates that may or may not represent independent species.

[b] Archaea.

known leviviruses are plasmid-specific coliphages that adsorb to F or sex pili. They have been divided, by serology and other criteria, into two genera. Several as yet unclassified leviviruses are specific for other plasmid types (e.g., C, H, M) or occur outside of the enterobacteria family (table 2-4).

Cystoviridae (dsRNA)

The family *Cystoviridae* consisted until recently of a single member, but several related viruses have recently been found. *Cystoviruses* are unique among bacteriophages because they contain three molecules of dsRNA and RNA polymerase. They have lipid-containing envelopes and have no morphological resemblance to other viruses with dsRNA, such as reoviruses or totiviruses (chapter 16).

Phages with Helical Symmetry

Inoviridae (ssDNA)

The family *Inoviridae* has two genera with very different host ranges. Their similarities in replication and morphogenesis seem to derive from the single-stranded nature of phage DNA rather than from a common origin. Despite the absence of lipids, virses are chloroform sensitive. The *Inovirus* genus includes 42 phages that are long, rigid, or flexible filaments of variable length and have been classified into 29 species by particle length, coat structure, and DNA content. They occur in a few Gram-negative bacteria and also in *Clostridium* and *Propionibacterium*. Viruses are sensitive to sonication and very resistant to heat. Many of them are plasmid-specific. The *Plectrovirus* genus includes 15 isolates. Phages are short, straight rods and occur in mycoplasmas only.

Lipothrixviridae (dsDNA)

The family *Lipothrixviridae* includes four viruses of the archaebacterial genus *Thermoproteus*. Particles are characterized by the combination of a lipoprotein envelope and rodlike shape.

Rudiviridae (dsDNA)

The family *Rudiviridae* includes two viruses of different length that have been isolated from thermophilic archaebacteria. Particles are straight, rigid rods without envelopes and closely resemble the tobacco mosaic virus.

Table 2-4 Host Range of Phages by Group and Host Genus

Phage group or family	Eubacteria	Archaea
Caudovirales	Any (ubiquitous)	Extreme halophiles and methanogens
Microviridae	Enterobacteria, *Bdellovibrio*, *Chlamydia*, *Spiroplasma*	
Corticoviridae	*Alteromonas*	
Tectiviridae	a. Enterics, *Acinetobacter*, *Pseudomonas*, *Thermus*, *Vibrio* b. *Bacillus*, *Alicyclobacillus*	
Leviviridae	Enterics, *Acinetobacter*, *Caulobacter*, *Pseudomonas*	
Cystoviridae	*Pseudomonas*	
Inoviridae *Inovirus*	Enterics, *Pseudomonas*, *Thermus*, *Vibrio*, *Xanthomonas*	
Plectrovirus	*Acholeplasma*, *Spiroplasma*	
Plasmaviridae	*Acholeplasma*	
Lipothrixviridae		*Acidianus*, *Sulfolobus*, *Thermoproteus*
Rudiviridae		*Sulfolobus*
Fuselloviridae		*Acidianus*, *Haloarcula*, *Sulfolobus*

Pleomorphic Phages

Plasmaviridae (dsDNA)

Only one certain member of the *Plasmaviridae* is known: *Acholeplasma* virus MVL2 or L2. It contains dsDNA, has no capsid, and may be called a nucleoprotein granule with a lipoprotein envelope. Four similar isolates are known, but one of them has been described as containing ssDNA and they cannot be classified.

Fuselloviridae (dsDNA)

The best-known member of the family *Fuselloviridae*, SSV1, persists in the archaeon *Sulfolobus shibatae* both as a plasmid and as an integrated prophage. It has been produced upon induction, but has not been propagated for absence of a suitable host. Particles are lemon-shaped with short spikes at one end. The coat consists of two hydrophobic proteins and host lipids. It is disrupted by chloroform.

"Guttavirus" (dsDNA)

The name "guttavirus" designates a droplet-shaped virus-like particle, named SDNV, which has been found in a *Sulfolobus* culture from a solfatara in New Zealand. It has received the status of an unassigned genus (14).

Taxonomical Physiology

Taxonomic subdivisions correlate with differences in physiology and life-style (2, 14) For example, most bacterial viruses infect bacteria from the outside after adsorption to the cell wall, capsules, pili, or flagella. However, cystovirus capsids, after losing their envelope, enter the space between cell wall and cytoplasmic membrane, while plasmaviruses infect their hosts by fusion of the viral envelope and mycoplasmal host cell membranes. Tailed phage DNA replication resembles somewhat that of herpesviruses (1) and is characterized by particulars which, to our knowledge, are not found in other phages. It generally starts at fixed sites of the DNA molecule, is bidirectional, and often generates giant DNA molecules, called concatemers, which are then cut to fit into phage heads. Gene expression is largely sequential. In phages with ssDNA, double-stranded replicative forms (RF) are produced. RNA molecules are never circularized. The RNA of leviviruses acts as mRNA and needs no transcription. *Microviridae*, *Leviviridae*, the *Inovirus* genus, and *Fuselloviridae* have overlapping genes that, by translation in different reading frames, allow the synthesis of several proteins from the same DNA or RNA segment.

In most phage families, the newly synthesized nucleic acid enters a preformed capsid; however, the levivirus capsid is constructed around or co-assembled with phage RNA. Assembly of tailed phages is a complex process and involves proteins acting in sequence and separate pathways for heads and tails, which are finally joined together. Tailed phage assembly often results in aberrant structures, for example giant or multi-tailed phages and tubules of polymerized head or tail protein, called polyheads, polytails, or polysheaths. Inoviruses frequently produce particles of abnormal length. The envelope of plasmaviruses is acquired by budding, but that of cystovirus ϕ6 is of cellular origin.

Further differences are seen in the mode of release. Lysis occurs in tailed and icosahedral phages and in the *Lipothrixviridae*. Bacterial cells are weakened from the inside and burst, liberating some 50 to 200 phages (sometimes many more). Inoviruses and fuselloviruses are liberated by extrusion, with phages being secreted through the membranes of their surviving hosts. Budding through bacterial membranes is found in plasmaviruses. Their host cells are not lysed and may produce phages for hours.

Lysogeny is widespread and not confined to tailed phages (1). Icosahedral phages are always virulent. About 50% of tailed phages are temperate. Lysogeny also occurs in the *Inovirus* genus and the *Lipothrixviridae*, *Plasmaviridae*, and *Fuselloviridae* families. Lysogeny is near-ubiquitous in the bacterial world and exists in eubacteria and in archaebacteria. Its frequency in a given bacterial species varies between 0 and 100% (often 40%) according to the species and quality of investigation. Tailed phages exhibit three types of lysogeny, exemplified by phages λ, P1, and Mu (chapter 7). The λ and P1 types seem to be equally frequent in tailed phages. The Mu type is exceptional in bacterial viruses. Remarkably, the integrase-mediated λ type also occurs in the *Fuselloviridae* and *Plasmaviridae* families (1). Phages of the genus *Inovirus* integrate into the host genome by means of host recombinases (7). The type of lysogeny in the *Lipothrixviridae* is unknown.

Host Range and Its Evolutionary Implications

Phages have been found in over 140 bacterial genera from all parts of the bacterial world: in archaebacteria and eubacteria, in aerobic, anaerobic, appendaged, budding, gliding, ramified, sporulating, or sheathed bacteria, in cyanobacteria, spirochetes, mycoplasmas, and chlamydias (table 2-5). Podovirus particles have even been observed in bacterial endosymbionts of paramecia. Indeed, bacteriophage hosts are represented in nearly all sections of *Bergey's Manual* (3). Tailed phages described in cultures of green algae and filamentous fungi are probably contaminants.

When phages are grouped according to phylogenetic bacterial groups established by rRNA sequencing (16), it becomes apparent that most major bacterial phyla are,

Table 2-5 Frequency of Phages in Bacterial Phylogenetic Divisions

Division or subdivision	*Myoviridae*	*Siphoviridae*	*Podoviridae*	Total of tailed phages	CFP
Archaea					
Euryarcheota	7	7		14	4
Crenarcheota					14[a]
Bacteria					
Bacteroides–Cytophaga –Flavobacterium	34	24	1	59	2
Chlamydiales					2
Cyanobacteria	22	6	16	44	
Deinococcus–Thermus	8	6		14	4
Firmicutes					
High G-C branch	3	482	21	506	
Low G-C branch	371	1300	86	1757	31
Fusobacteria		1	4	5	
Proteobacteria: all	757	729	536	2022	128?
γ subdivision only	642	530	429	1601	111?
Spirochetes	10	1		11	
Total	1212	2555	660	4427	186?

Based on the organism list of the National Center for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov/taxonomy/) (15) and on phage counts from (3). Total phage numbers are lower than in table 2-1 because rRNA data are unavailable for some phage hosts.
CFP, cubic, filamentous, and pleomorphic phages.
[a] Including a phage-like particle.

so far, without phages. Most phages are found in easily grown and medically or industrially important bacteria:

1. *Firmicutes* with high G-C (coryneforms, mycobacteria, streptomycetes),
2. *Firmicutes* with low G-C (bacilli, lactobacilli, lactococci, clostridia, staphylococci, streptococci),
3. Proteobacteria, especially of the γ subdivision. The latter includes enterobacteria (over 800 phage observations) and pseudomonads.

Tailed phages predominate almost everywhere and occur in both eubacteria and *Euryarcheota*, suggesting that they originated before separation of these bacterial kingdoms. *Siphoviridae* are particularly frequent in actinomycetes, coryneforms, lactococci, and streptococci. Myoviruses and podoviruses are relatively frequent in enterobacteria, pseudomonads, bacilli, and clostridia. This distribution is probably related to features of bacterial speciation, for example to evolution of particular receptors or restriction endonucleases.

There is a chasm between the ubiquitous tailed phages and the other types of bacterial viruses. The latter are relatively rare and have narrow, particular host range. Inoviruses of the *Plectrovirus* genus are restricted to mycoplasmas, and *Fuselloviridae*, *Lipothrixviridae*, and *Rudiviridae* are limited to a particular archaeal subdivision, the extremely thermophilic *Crenarcheota*. This suggests that their hosts constitute ecological niches, in which these viruses originated. On the other hand, there are oddities in distribution: (i) filamentous inoviruses, though generally associated with enterobacteria and their relatives, occur in such diverse bacteria as *Clostridium*, *Propionibacterium*, and *Thermus*; and (ii) tectiviruses are found in enterics, bacilli, and *Thermus*. As these phages are plasmid-dependent, their distribution may be explained by plasmid transfer (3).

Why Phage Classification?

To microbiologists focusing on a single microorganism, classification may appear a sterile exercise. Others, impressed by spectacular advances in limited fields, may consider that "we know it all." Indeed, bacteriophages T4 and λ are among the best-known viruses of all. This way of thinking is short-sighted and self-defeating. We have barely scratched the surface of virology. With respect to bacteriophages, (i) research has concentrated on a few viruses and neglected the 5000 others known, (ii) research is limited to about 15 countries in the world, and (iii) important habitats such as hot springs or fermentors have been little explored. My experience is that, armed with an electron microscope, I am certain to find scores of new phages in a bottle of sewage. At the same time, phages have practical applications that demand precise identification. Last but not least, microbiologists want to understand the living world and many of them are actively engaged in teaching. For them, and their students, taxonomy is simplification. It is impossible and pointless to memorize the properties of 5000 individual tailed phages, but it is much more rewarding to study tailed phages as a group. The advantages of classification, especially with respect to phages, are many:

1. Classification is generalization and simplification.
    a. It promotes comparison and thus virus research and better understanding of the viral world.

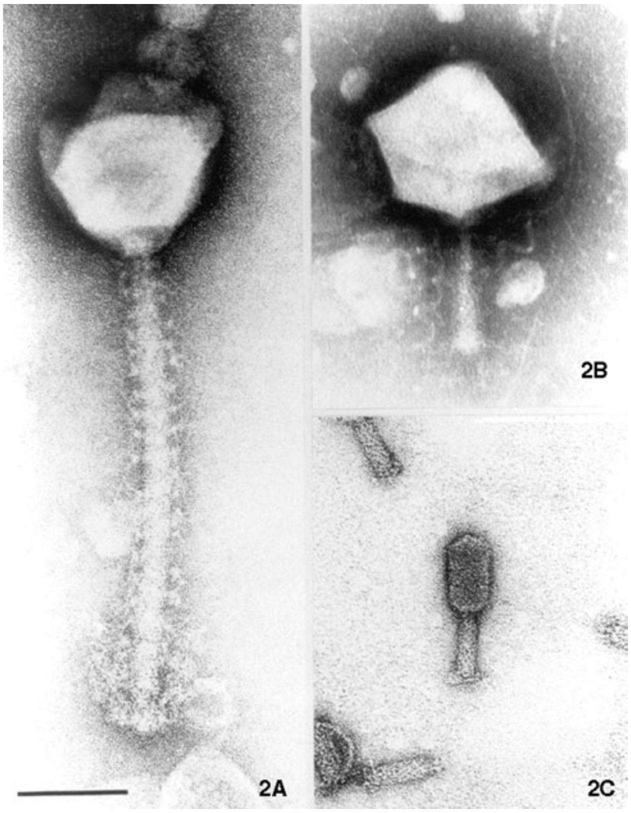


Figure 2-2 Selected myoviruses. A: *Bacillus megaterium* phage G, the largest myovirus known; note the spiral filament around the tail. B: Giant unknown bacteriophage found in macerated debris of *Bombyx mori*; note the wavy tail fibers. Phage heads of this size are easily deformed. C: *Bdellovibrio bacteriovorus* phage ϕ1402, the smallest myovirus known (isolated by B. A. Fane, Department of Veterinary Science, University of Arizona, Tucson, AZ). × 297,000; bar indicates 100 nm. Phosphotungstate (2%, pH 7.2) (A, B) and uranyl acetate (2%, pH 4.0) (C).

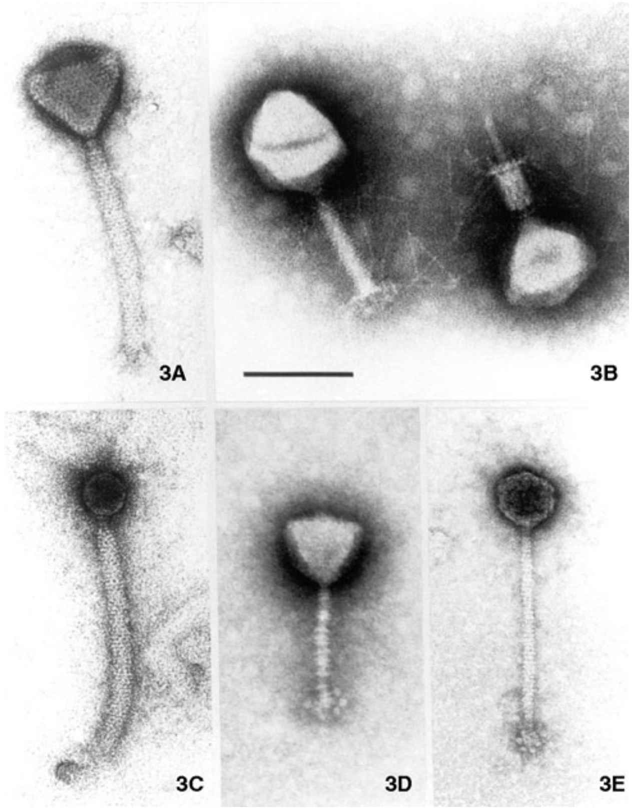


Figure 2-3 Myoviruses (A–C) and siphoviruses (D, E). A: *Staphylococcus hyicus* phage Twort. B: Coliphage RB69. The head of the left particle is slightly deformed and the tail of the right particle is contracted. C: "Killer particle" of *Bacillus subtilis*; note the small, bacterial DNA-containing head. D: Phage NM1 of *Sinorhizobium meliloti* with transverse tail disks and globular fixation structures. E: *Bacillus subtilis* phage ϕ105. × 297,000; bar indicates 100 nm. Uranyl acetate (A, C, E) and phosphotungstate (B, D).

   b. It is indispensable for teaching. No textbook is conceivable without it and students, as teacher well know, want certitude. Taxonomy is also a basic ingredient in theses. Without classification, virology is just a magma for the beginner.

2. Classification is essential for phylogenetic studies. It identifies the very biological groups that are to be compared. For example, isolated genes or genome sequences are of limited interest, but gain immensely in significance when traced to other categories of organisms. Without phage classification, we would probably not know that such things as horizontal gene transfer exist.

3. Classification is required for proper identification of:

   a. New phages.
   b. Harmful phages in biotechnology and industrial fermentations.
   c. Industrially important phages in patent applications.
   d. Therapeutic phages. Now that phage therapy is making a comeback, they must be identified. It is indeed inconceivable that people should be treated (even injected) with something unknown.

4. Classification is a great research aid. For example, the DNA size of tailed phages can be predicted with accuracy from capsid dimensions.

Outlook

Phage taxonomy is bound to evolve. New phage families are likely to be discovered in unusual habitats, such as volcanic springs, hypersaline lagoons, or the mammalian

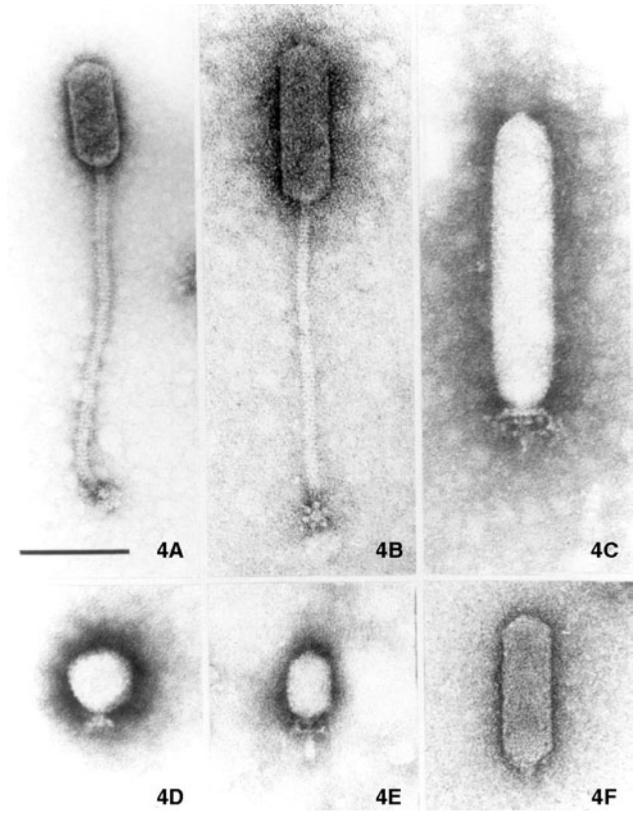


Figure 2-4 Siphoviruses (A, B) and podoviruses (C–F). A: *Staphylococcus aureus* phage 6. B: *Bacillus subtilis* phage BS5. C: *Lactococcus lactis* ssp. *cremoris* phage KSY1. D: *Salmonella typhimurium* phage P22. E: *Bacillus* sp. phage GA-1. F: Coliphage Esc-7-11. × 297,000; bar indicates 100 nm. Uranyl acetate (A, B, F) and phosphotungstate (C, D, E).

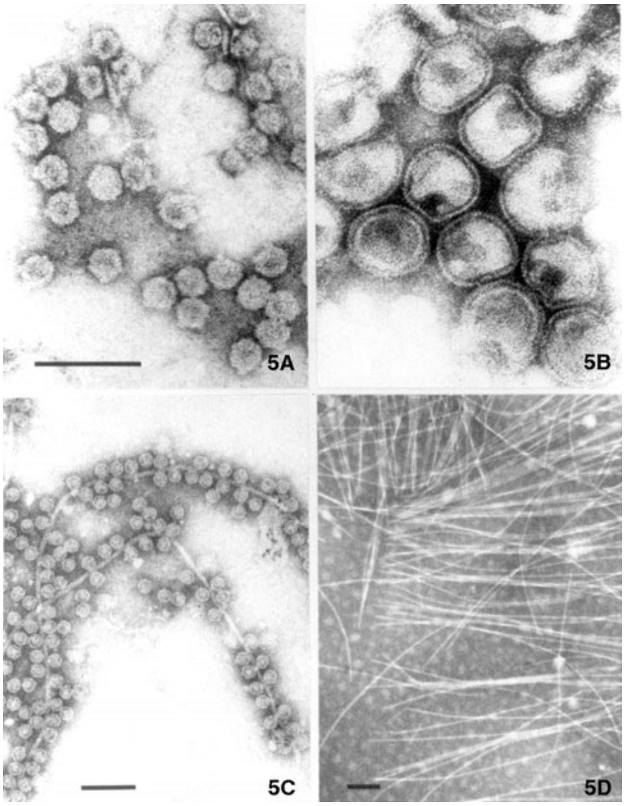


Figure 2-5 Selected isometric and filamentous phages. A: *E. coli* microvirus ϕX174, × 297,000. B: Tectivirus 37–64 of *Thermus* sp.; note inner membrane and deformed particles; × 297,000 C: *E. coli* levivirus R17 adsorbed to pili, × 148,500 D: Inovirus H75 of *Thermus thermophilus*, × 92,400. Bars indicate 100 nm. Uranyl acetate (A) and phosphotungstate (B–D).

rumen. Indeed, strange phage-like particles that probably represent novel families have been reported many times, for example arrow-shaped, apparently archaebacteria-related particles in saline environments (18). Families are the most stable part of the present edifice of viral taxonomy. Genera and species are more fluid because there are no universally accepted or applicable criteria for defining them. Mayr's "biological species definition" (11), based on interbreeding with production of fertile offspring, was devised for songbirds and is almost inapplicable to viruses. The polythetic species concept (13), which is in principle applicable to genera, is a concept and does not provide concrete guidelines for definitions of phage species or genera. I contend that no biologist (or virologist) can certify what a species is. Virus (or phage) classification is therefore still an art.

The most important present challenge is the interpretation of horizontal gene transfer in tailed phages. It may be argued that any classification of these viruses is impossible because of gene shuffling and the occurrence of certain genes in apparently unrelated viruses. This attitude is a dead end which can only lead to forfeiting the benefits of classification. It overlooks major realities of viral biology: (i) Common genes do not necessarily indicate close relationships. Horizontal gene transfer is frequent and ubiquitous. Some (all?) genes move through the living world; for example T4 lysozyme surfaces in goose eggs and human tears (1). (ii) Recombination can mask or simulate relationships. (iii) In viruses related by evolution, genomic relationships between different taxa are to be expected. The question of devising genera or species in tailed phages thus becomes a quantitative one. We may have to follow the bacteriologists who, arbitrarily, fixed the threshold value for species delineation at 60–70% DNA homology (8).

References

1. Ackermann, H.-W. 1998. Tailed bacteriophages—The order *Caudovirales*. Adv. Virus Res. 51:135–201.

2. Ackermann, H.-W. 1999. Bacteriophages, pp. 398–411. *In* J. P. Lederberg (ed.-in-chief), Encylopedia of Microbiology, 2nd edn, vol. 1. Academic Press, New York.
3. Ackermann, H.-W. 2001 Frequency of morphological phage descriptions in the year 2000. Arch. Virol. 146: 843–857.
4. Bradley, D. E. 1967. Ultrastructure of bacteriophages and bacteriocins. J. Bacteriol. 31:230–314.
5. D'Hérelle, F. 1918. Technique de la recherche du microbe filtrant bacteriophage *(Bacteriophagum intestinale)*. C. R. Soc. Biol. 81:1160–1162.
6. Duckworth, D. H. 1976. Who discovered bacteriophage? Bacteriol. Rev. 40:793–802.
7. Huber, K. E., and M. K. Waldor. 2000. Filamentous phage integration requires the host recombinases XerC and XerD. Nature 417:656–659.
8. Johnson, J. L. 1984. Nucleic acids in bacterial classification, p. 8–11. *In* N. T. Krieg and J. G. Holt (eds.), Bergey's Manual of Systematic Bacteriology, vol. 1. Williams & Wilkins, Baltimore, Md.
9. Lwoff, A., R. W. Horne, and P. Tournier. 1962. A system of viruses. Cold Spring Harbor Symp. Quant. Biol. 27:51–62.
10. Maniloff, J., and H.-W. Ackermann. 1998. Taxonomy of bacterial viruses: establishment of tailed virus genera and the order *Caudovirales*. Arch. Virol. 143:2051–2063.
11. Mayr, E. 1942. Systematics and the Origin of Species from the Viewpoint of a Zoologist, p. 119. Columbia University Press, New York.
12. Skalka, A., and P. Hanson. 1972. Comparisons of nucleotides and common sequences in deoxyribonucleic acid from selected bacteriophages. J. Virol. 9:583–593.
13. Van Regenmortel, M. H. V. 1990. Virus species, a much overlooked, but essential concept in virus classification. Intervirology 31:241–254.
14. Van Regenmortel, M. H. V., C. M. Fauquet, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, J. Maniloff, D. J. McGeoch, C. R. Pringle, and R. B. Wickner (eds.). 2000. Virus Taxonomy, pp. 63–136, 267–284, 389–393, 645–650. Seventh Report of the International Committee on Taxonomy of Viruses. Academic Press, San Diego, Calif.
15. Vander Byl, C., and A. M. Kropinski. 2000. Sequence of the genome of *Salmonella* bacteriophage P22. J. Bacteriol. 182:6472–6481.
16. Wheeler, D. L., C. Chappey, A. E. Lash, D. D. Leipe, T. L. Madden, G. D. Schuler, T. A. Tatusova, and B. A. Rapp. 2000. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 28:10–14.
17. Wildy, P. 1971. Classification and nomenclature of viruses. First Report of the International Committee on Nomenclature of Viruses. Monographs in Virology, vol. 5. Karger, Basel.
18. Zillig, W., A. Kletzin, C. Schleper, I. Holz, D. Janekovic, J. Hain, M. Lanzendörfer, and J. K. Kristjansson. 1994. Screening for *Sulfolobales*, their plasmids, and their viruses in Icelandic solfataras. Syst. Appl. Microbiol. 16:609–628.

3

Prophage Genomics

HARALD BRÜSSOW

At the time of this writing, the GenBank phage database comprises 200 complete phage genome sequences. An equivalent number of phage sequences were passively acquired as prophages in bacterial genome sequencing projects. The scientific value of these prophage sequences was only recently recognized (13, 14). It goes beyond their potential to double the content of the current phage database and to correct a bias of the database towards selected phage systems (coliphages, dairy phages, mycobacteriophages) (7, 9). These prophage sequences allowed a first insight into the evolution of phage–host genome interactions at a molecular level. This analysis turned out to be especially fruitful for bacterial pathogens (1, 5, 49).

Technical Difficulties

Few phage genes are sufficiently conserved and distinct from bacterial genes to serve as markers for prophage sequences. For temperate *Siphoviridae*, suitable phage proteins comprise the phage integrase (10), the portal protein, the terminase (17), and the tail tape measure protein. A computer program (25) combining semantic searches of the gene annotations with BLAST searches identified automatically most of the prophages compiled by Casjens in a more labor-intensive approach (14). However, other types of phages can integrate their DNA into the bacterial chromosome: for example, P2- and Mu-like *Myoviridae*, *Inoviridae* in *Vibrio* and *Xanthomonas*, *Plasmaviridae* in *Acheloplasma*. In addition to psiM1-like *Siphoviridae*, Lipothrix- and Fuselloviruses integrate their genomes into the chromosomes of Archaea. Still other forms of lysogeny exist that do not lead to the integration of phage DNA into the bacterial chromosome: for example, prophages P1 and N15 are maintained as circular or linear plasmids, respectively. However, different groups of temperate phages show a remarkable conservation of their gene map (synteny) that allowed a tentative identification of prophage sequences with only a few landmark database matches (8, 16, 35). A final difficulty with the detection of prophage sequences in bacterial genomes is the trend for progressive deletion and rearrangements of prophage DNA (18, 31).

Distribution of Prophage Sequences

When published bacterial genomes were screened for prophage sequences, approximately half of them scored positive (13, 14, 31). A clear bias was observed in the distribution: most Archaea (only one subgroup showed phagelike viruses) and intracellular bacterial parasites lacked prophages, while many bacterial pathogens of humans, animals, and plants frequently showed a high prophage content. When all prophage sequences were located on an idealized circular bacterial genome map, no preferred positions of the prophages was observed. However, when individual bacterial groups were investigated, clear biases in the prophage distribution were observed. For example, prophages from *Escherichia coli* showed a trend for location in the first half of the chromosome (i.e., between the origin and the terminus of bacterial replication), while prophages from *Streptococcus pyogenes* tended to cluster at both sides of the terminus of replication (13). Interestingly, prophages from low G-C content, Gram-positive bacteria were differently oriented when located to the left or the right of the terminus of bacterial replication. The lytic gene cluster always pointed in the direction of the majority of the surrounding bacterial genes.

In highly prophage-infested bacterial genomes such as *E. coli* O157 (37), *Lactococcus lactis* (15), and *Streptococcus pyogenes* (3, 22, 41) up to 8% of the bacterial chromosome consisted of prophage sequences. Where multiple strains of the same bacterial species were sequenced, prophage DNA accounted for a substantial amount of inter-strain genetic variability. For example, genomic comparison between the pathogenic *E. coli* strain O157 EDL933 and the laboratory *E. coli* strain K-12 revealed 4.1 Mb of common chromosome backbone sequence, 1.3 Mb of O157-specific, DNA and 0.5 Mb of K12-specific DNA (37). Approximately half of the O157-specific DNA was prophage DNA. Even more extreme cases are known: when genomes from different M serotypes of *S. pyogenes* were compared the major gaps in the dot-plot DNA sequence alignment were nearly exclusively prophage sequences. Similarly, DNA–DNA hybridizations in the microarray format revealed that prophages frequently made important contributions to the strain-specific gene complement of the sequenced reference strain. This varied from less than 10% in bacteria containing only one or a few prophage remnants, as in the currently sequenced high G-C content, Gram-positive bacteria (*Mycobacterium, Bifidobacterium*), to major contributions in low GC content, Gram-positive bacteria, where the percentage ranged from about 30% in *S. agalactiae* (2 prophages, (43)) to about 50% in *Lactobacillus johnsonii* (2 prophages, (19, 46)) and up to 90% in *S. pyogenes* strains containing three to six prophages (41).

Theoretical Framework

The peculiar life-style of temperate phages makes them model systems for addressing a number of fundamental questions in evolutionary biology. The viral DNA undergoes different selective pressures when replicated as phage DNA during lytic infection cycles than it does as prophage DNA maintained in the bacterial genome during lysogeny. Darwinian considerations along the lines of the selfish gene concept lead to interesting conjectures (9, 13, 18, 31). One could anticipate that the prophage decreases the fitness of its lysogenic host by at least two processes: first by the metabolic burden of replicating extra DNA and second by the lysis of the host after prophage induction. To compensate for these disadvantages one has to hypothesize that temperate phages encode functions that increase the fitness of the lysogen. According to the selective value of these phage genes, the lysogenic cell will be maintained or even be over represented in the bacterial population. An obvious selective advantage for the lysogenic host is the immunity and superinfection exclusion genes of the prophage that protect the lysogen against phage infection. These genes are also of direct advantage to the prophage since they exclude foreign phage DNA from competing with the resident prophage DNA for the same host. Where phages from the environment are not a sufficiently strong selection pressure, other phage genes have to increase the fitness of the lysogenic host, frequently in rather unanticipated ways (lysogenic conversion genes). Classic examples of such phage-encoded genes that increase host fitness include diphtheria toxin, streptococcal erythrogenic toxin A, and the nonessential phage λ gene *bor* that confers serum-resistance to the *E. coli* lysogen (2). Interestingly, lysogenic *E. coli* clones were also more competitive than prophage-free clones in laboratory growth (21). In these cases, the reproductive success of the lysogenic bacterium translates directly into an evolutionary success for the resident prophage.

However, host–parasite relationships are also an arms race and represent therefore a highly dynamic genetic equilibrium. Gains from prophages carrying genes that increase host fitness are short-lived from a bacterial standpoint if the resident prophage ultimately destroys the bacterial lineage. In this way, prophages can be considered to be dangerous molecular time bombs that can kill the lysogenic cell upon their eventual induction (31). One would therefore expect evolution to select lysogenic bacteria with mutations in the prophage DNA. Mutations that inactivate the prophage induction process avoid the loss of the lysogenic clone from the bacterial population. In a next step, one would expect that selection pressures lead to large-scale deletion of prophage DNA in order to decrease the metabolic burden of extra DNA synthesis. One predicts furthermore that useful prophage genes (e.g., lysogenic conversion genes) are preferentially spared from this deletion process since their loss would actually decrease the fitness of the cell. It was proposed that a high genomic deletion rate is instrumental in removing dangerous genetic parasites from the bacterial genome. Deletion processes could explain why the bacterial genomes did not increase in size despite a constant bombardment with parasitic DNA over evolutionary time periods. The streamlined bacterial chromosome containing few pseudogenes might be the consequence of this deletion process of parasitic DNA (31).

Prophages in *Streptococcus Pyogenes*

An interesting test case for the predictions of the theoretical model is the important human pathogen *S. pyogenes*. The invasive *S. pyogenes* M1 strain SF370 contains eight prophage elements (18, 22) (figure 3-1). Only prophage SF370.1 could be induced by mitomycin C treatment. This prophage is a *pac*-site temperate member of the *Siphoviridae* found in many different lactic acid bacteria. Its closest relative was prophage NIH1.1 from a Japanese M3 *S. pyogenes* strain (12, 27). Notably, possession of prophage NIH1.1 differentiated older from newly emerging *S. pyogenes* strains in Japan (28). Prophage acquisition might thus be a major mechanism of short-term evolution in this epidemiologically highly dynamic bacterial species. In fact, extensive genomics analysis combined with epidemiological surveys of *S. pyogenes* isolates over the last decades by clinical microbiologists led to an attractive model. James Musser and colleagues suggested that the recently emerged unusual virulent subclones of M3 *S. pyogenes* strains are the result of sequential acquisition of three prophages and their associated virulence genes over the last 80 years (1, 3). If this observation can be generalized to other bacterial pathogens, medical microbiologists have to confront foes that evolve in the fast lane. Like in the acquisition of antibiotic resistance genes, the development of bacterial pathogenicity seems to rely heavily on the use of mobile DNA elements. It is

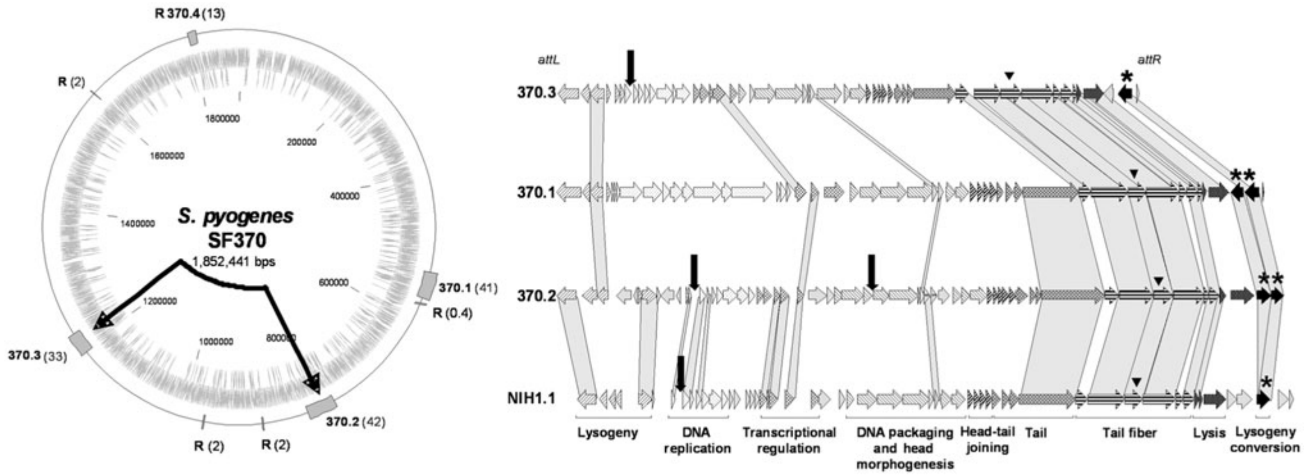


Figure 3-1 Location of *Streptococcus pyogenes* strain SF370 prophage and their genomic maps. Left: Location and relative size (in kilobases) of prophages and prophage remnants R on the *S. pyogenes* strain SF370 genome map. The linked arrows indicate a possible site for homologous recombination between closely related prophage DNA segments. Right: The genome maps of three SF370 prophages and a prophage from *S. pyogenes* strain NIH1 are aligned with the attachment sites at the left and right ends. Prophage NIH1.1 has been identified as a genetic marker for recently emerged clinical isolates of *S. pyogenes* in Japan. The phage modules are: lysogeny, DNA replication, probable transcriptional regulation, DNA packaging and head, morphoggenesis, head-to-tail joining, tail synthesis, tail fiber synthesis and assembly, lysis, and lysogenic conversion genes encoding superantigen/mitogenic factors. Large vertical arrows indicate mutations that are likely to inactivate the prophage. An asterisk marks phage genes that potentially contribute to the virulence of the lysogenic host. The phage hyaluronidase is labeled by a triangle. Regions of DNA sequence similarity between the prophages are connected by shading (for details see 13). See thebacteriophages.org/frames_0030.htm for a color version of the figure.

currently unclear why resistance genes come with plasmids and transposons, while virulence genes show a tendency to travel with the genomes of temperate phages.

All 20 complete prophages detected in the currently sequenced *S. pyogenes* strains showed likely lysogenic conversion genes between the phage lysin gene and the right attachment site. Prophage SF370.1 encoded the pyrogenic exotoxin C and a mitogenic factor (22). Related proteins were encoded in the other *S. pyogenes* prophages covering distinct members of superantigens, mitogenic factors (DNases), and toxic enzymes. These phage proteins may contribute to the immune deregulation observed during invasive streptococcal infections. The lysogenic conversion genes in the prophages differ in their G-C content from the surrounding prophage and bacterial DNA (22). Their location in the vicinity of the phage attachment site suggested a faulty phage excision process in an unusual bacterial host of lower G-C content as the origin of this DNA. The subsequent spread of these genes is also suggested by the presence of sequence-identical genes in the veterinary pathogen *S. equi*. Conceptually, bacteria that can acquire multiple prophages with related but distinct forms of superantigens or other virulence factors can play a type of combinatorial biology for the construction of a successful pathogen.

Streptococcus pyogenes showed other elements predicted by theory, for example prophage inactivation. Prophage SF370.3 showed a 33 kb-long genome that closely resembled the genome organization of the *cos*-site temperate *Lactococcus lactis* phage r1t (18). Analysis of the prophage genome revealed mutations in the replisome organizer gene that may prevent the induction of the prophage. Prophage SF370.2 showed a 43 kb-long genome that again resembled the genome organization of *pac*-site temperate *Siphoviridae* infecting dairy bacteria. SF370.2 showed two inactivating mutations (one in the replisome organizer gene and another in the gene encoding the portal protein) and four DNA insertions into the prophage DNA (18) (figure 3-2).

A clear trend for prophage genome reduction was documented by the many *S. pyogenes* prophage remnants that are probably the result of massive losses of prophage DNA (18). The largest prophage remnant, SF370.4, showed a 13 kb-long genome consisting of lysogeny, DNA replication, and transcriptional regulation genes still flanked by attachment sites. Other prophage remnants were much smaller and consisted frequently only of the phage integrase and one or a few associated prophage genes (a *cI*-like phage repressor was a frequent finding).

In *S. pyogenes*, prophages are apparently hotspots for genetic recombination. A particularly conserved genome region is the tail fiber gene cluster, especially around the hyaluronidase gene. This tail fiber protein must provide the phage with access to the target cell through the hyaluronic-acid-containing bacterial capsule. Some post-streptococcal patients mount antibodies against this phage protein. Since hyaluronic acid is also

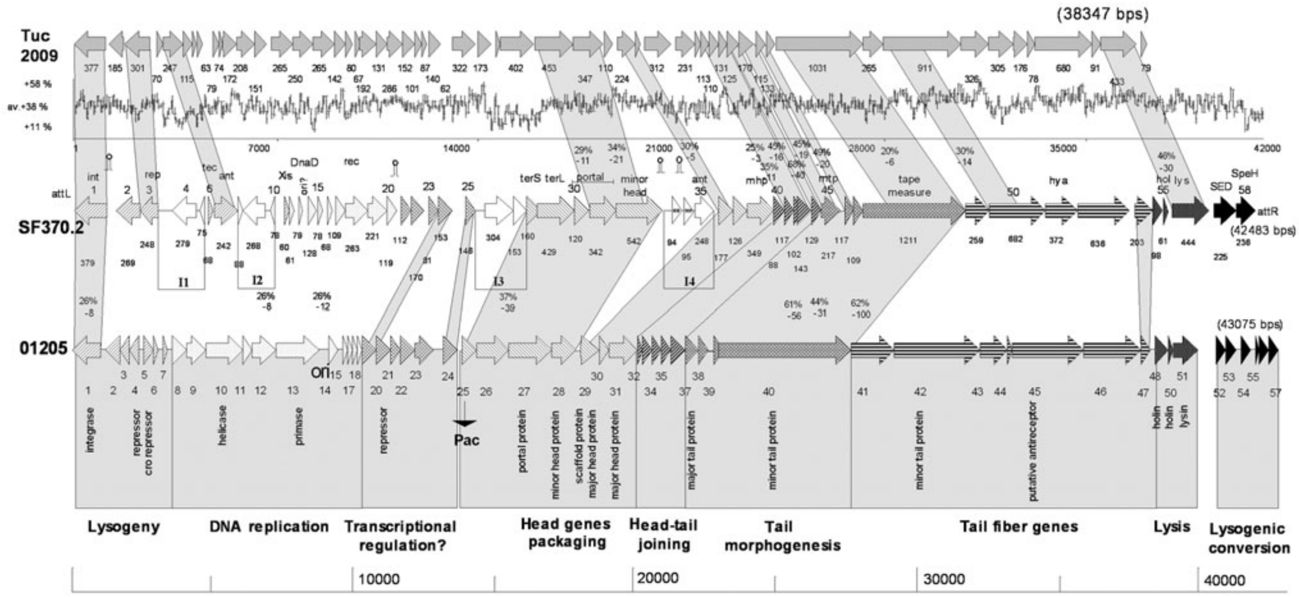

Tuc 2009
(38347 bps)
+58 %
av.+38 %
+11 %
attL
SF370.2
(42483 bps)
attR
01205
(43075 bps)
int
rep
tec
ant
DnaD
Xis
ori?
rec
terS terL
portal
minor head
ant
mhp
mtp
tape measure
hya
hol lys
SED
SpeH
I1
I2
I3
I4
ori
Pac
integrase
repressor
cro repressor
helicase
primase
repressor
portal protein
minor head protein
scaffold protein
major head protein
major head protein
major tail protein
minor tail protein
minor tail protein
putative antireceptor
holin
holin
lysin
Lysogeny
DNA replication
Transcriptional regulation?
Head genes packaging
Head-tail joining
Tail morphogenesis
Tail fiber genes
Lysis
Lysogenic conversion
10000
20000
30000
40000

Figure 3-2 Alignment of the phage genome maps. Top: *Lactococcus lactis* phage Tuc2009. Middle: *Streptococcus pyogenes* prophage SF370.2. Bottom: *Streptococcus thermophilus* phage O1205. Open reading frames were depicted that exceeded 60 codons in length and that started with an ATG codon (exceptions are ORFs 33 and 34 which start with GTG). The genes were identified by their codon length or were numbered starting with the integrase gene. Selected proteins were annotated with their putative functions. The genes are connected with shading when the predicted proteins shared significant sequence similarity. Selected similarities are documented by percent amino acid identity and their log BLAST E-values. Putative insertions are noted I1 to I4. Possible rho-independent terminators are indicated with a hairpin. The wavy line provides the GC content curve for SF370.2 over the genome scale in base pairs. See thebacteriophages.org/frames_0030.htm for a color version of this figure.

part of the connective tissue, this immune response seems to suggest that this prophage protein might assist in the propagation of the pathogen along tissue planes of patients.

The conserved DNA segments around the hyaluronidase gene and short conserved DNA segments near the *attR* sites might be targets for homologous recombination between the prophage and a superinfecting phage and thus result in the reshuffling of virulence genes between phages. However, the region around the phage hyaluronidase genes might serve also as target sites for homologous recombination between two prophages residing in the same host chromosome. This type of recombination will then result in rearrangements of the bacterial genome (9). The genome comparison of a US and a Japanese M3 serotype *S. pyogenes* strain provided convincing evidence for such a prophage-mediated bacterial genome rearrangement event (36).

Prophages in the Genomes of Dairy Bacteria and Gut Commensals

Lactococcus lactis is a major bacterial starter organism used in industrial cheese fermentation. Strain IL1403 contains in its 2.3 Mb genome six prophage elements; all were flanked by the phage attachment sites and with one exception could be excised from the bacterial genome (15). However, only two prophages gave rise to infectious phage particles after prophage induction. Two prophages resembled in their genome organization the temperate *cos*-site *L. lactis* phage BK5-T, an Sfi21-like siphovirus. In the lysogen, only two genome regions of prophage BK5-T were transcribed (4): one transcript started in the *cI*-like repressor gene extending into the lysogeny module, while another transcript was located between the lysin gene and the right attachment site, the region where *S. pyogenes* prophages encode virulence factors. Three 15 kb-long DNA segments in IL1403 represent likely prophage remnants. All had conserved part of the lysogeny module (integrase/repressor) and, in variable amounts, DNA replication and a few structural genes (15).

Many temperate *Siphoviridae* from a wide range of low GC content, Gram-positive bacteria showed extra genes between the lysin gene and *attR*. Prophages from the pathogen *Staphylococcus aureus* encoded virulence genes near *attR* (leukotoxins, enterotoxins, exfoliative toxins, staphylokinase) (13). Two different types of prophages from *Streptococcus thermophilus*, the starter used for yogurt fermentation, showed genes without database matches in that region. Transcription analysis demonstrated that these phage genes were the only prominently expressed prophage genes in addition to the *cI*-like repressor and the superinfection exclusion genes (45). The rest of the prophage genome is

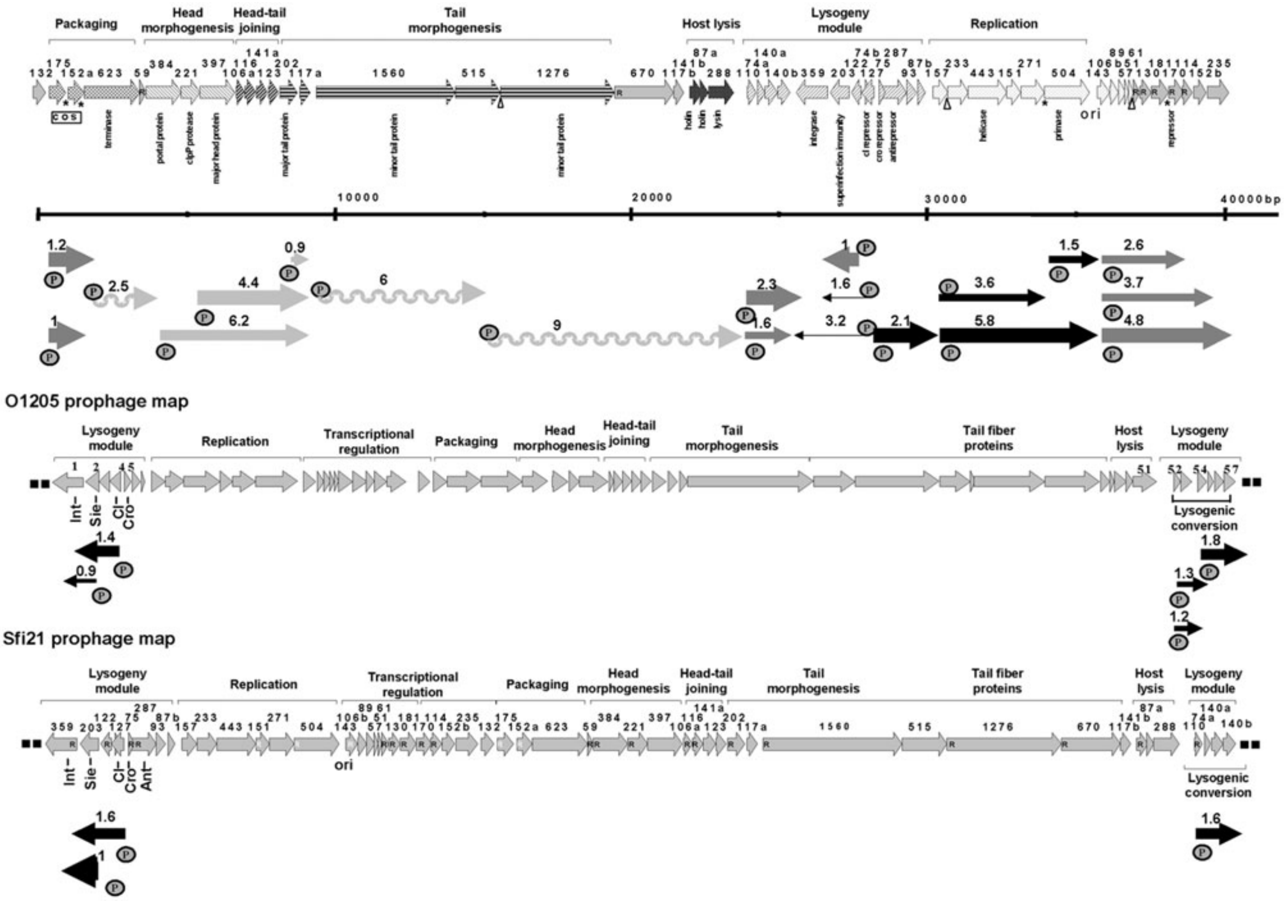

Packaging
Head morphogenesis
Head-tail joining
Tail morphogenesis
Host lysis
Lysogeny module
Replication
cos
terminase
portal protein
clpP protease
major head protein
major tail protein
minor tail protein
minor tail protein
holin
holin
lysin
integrase
superinfection immunity
cI repressor
cro repressor
antirepressor
helicase
primase
ori
repressor
10000
20000
30000
40000bp
1.2
2.5
4.4
0.9
6
1
6.2
9
2.3
1.6
1
1.6
3.2
2.1
3.6
5.8
1.5
2.6
3.7
4.8
O1205 prophage map
Lysogeny module
Replication
Transcriptional regulation
Packaging
Head morphogenesis
Head-tail joining
Tail morphogenesis
Tail fiber proteins
Host lysis
Lysogeny module
Int
Sie
CI
Cro
Lysogenic conversion
1.4
0.9
1.8
1.3
1.2
Sfi21 prophage map
Lysogeny module
Replication
Transcriptional regulation
Packaging
Head morphogenesis
Head-tail joining
Tail morphogenesis
Tail fiber proteins
Host lysis
Lysogeny module
Int
Sie
CI
Cro
Ant
ori
Lysogenic conversion
1.6
1
1.6

Figure 3-3 Transcription maps of temperate *Streptococcus thermophilus* phages. Top: Genome map of phage Sfi21 as found in the extracellular phage particle. The individual phage modules as determined by bioinformatic analysis are marked with the brackets above the map. The arrows below the map indicate the transcripts: early, middle, and late. The mRNA length in kilobases is indicated by the numbers above the arrows. The 5′ start sites, determined by primer extension analysis, are marked by a P in a circle. The width of the arrow indicates the strength of the autoradiographic signal as determined by Northern blots. The wavy lines indicate smeared hybridization results with a sharp upper end (48). Bottom: Genome maps of *S. thermophilus* prophages O1205 and Sfi21 as they are integrated in the bacterial chromosome. The arrows under the map mark the transcribed prophage genes (45). Note that the transcription of the prophages is limited to the regions near both attachment sites. See thebacteriophages.org/frames_0030.htm for a color version of this figure.

transcriptionally silent, well in contrast to lytic infections engendered by the same phages. Notably, these candidate lysogenic conversion genes belonged to the early genes transcribed in the lytic infection cycle (48) (figure 3-3). However, database matches provided no hints for a potential function of these genes. Interestingly, *S. thermophilus* prophage TP-J34 that integrates into a tRNA gene without disrupting it, showed a lysogenic conversion phenotype, that is a change in growth property (aggregated vs. planktonic growth) (H. Neve, personal communication). A prophage remnant was also detected in *S. thermophilus* (45). The phage attachment sites flanked the phage integrase and two nonattributed but transcribed prophage genes, suggesting a sparing of potential lysogenic conversion genes from the deletion process as predicted by the theoretical framework.

The prophages from lactobacilli, a third group of dairy bacteria and important commensals of humans and animals, were investigated in some detail (see chapter 41). The prophages from a *Lactobacillus plantarum* strain are instructive in the present context (47). Two prophage regions contained transcribed extra genes. They were located between the phage lysin and *attR* and between the phage repressor and the integrase genes. Notably, several of these extra genes shared 30–40% protein sequence identity with candidate virulence genes from bacterial pathogens, most prominently with candidate lysogenic conversion genes from *S. pyogenes* prophages (47). In addition, *Lactobacillus* prophages contained transcribed tRNA genes.

Prophages in *E. coli*

The laboratory *E. coli* strain K-12 contains, in addition to phage λ, eight prophage remnants (11). Three are

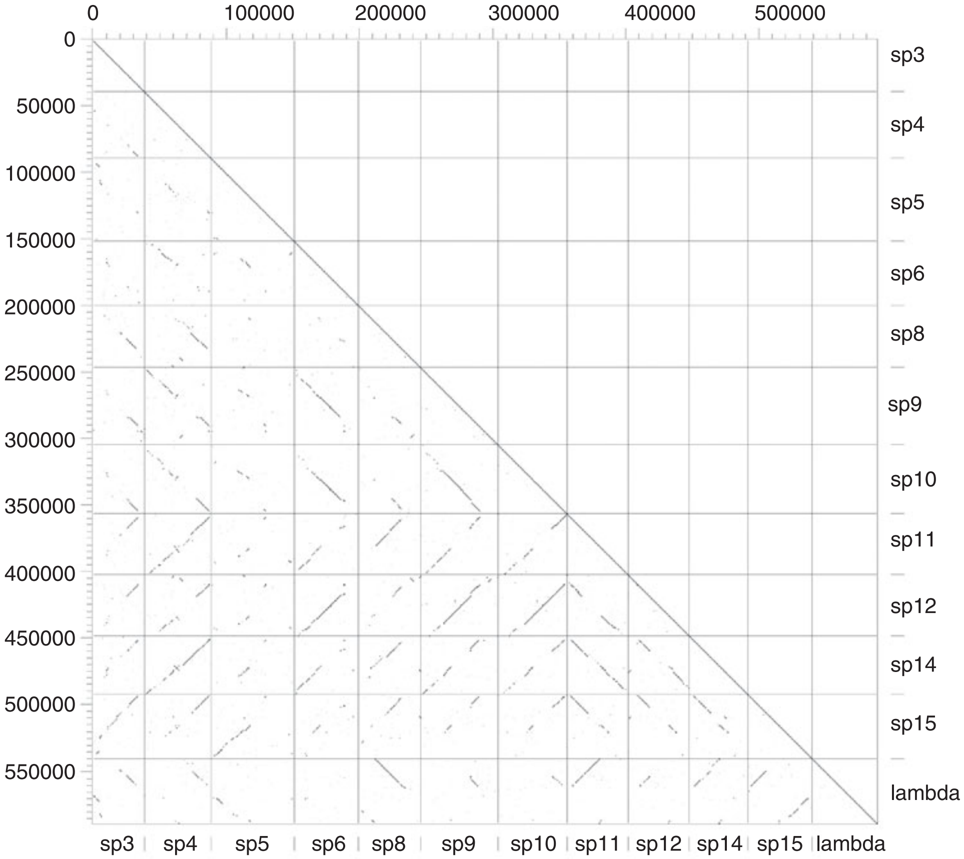


Figure 3-4 Dot-plot matrix for phage λ and the 11 λ-like prophage elements. Prophage elements, Sp3 to Sp15 from the *E. coli* strain O157:H7 Sakai. Phage λ is given as the conventional vegetative map (starting with the *cos*-site and the packaging genes; chapter 27) while the Sakai prophages start with *attL* and the integrase gene. All prophages except Sp6 share DNA sequence identity with the genes of λ. Sp6 is clearly distinct from λ at the DNA level while it shares along regions of DNA identity with other Sakai prophages. Remarkable is the high degree of DNA sequence identity between the different Sakai prophages. The sequence identity should lead to frequent genome rearrangements due to homologous recombination between the prophage elements.

sequence-related to prophages integrated at analogous loci in the pathogenic O157 *E. coli* strains (26, 39). However, 933R prophage in O157 is far more complete than its Rac homolog in K-12 and apparently misses only the *N-kil* gene region. *Shigella flexneri*, which separated from K-12 less than 1 million years ago (29), can be aligned with K-12 at the DNA level. Some of the gaps in the alignment were made up by many likely prophage remnants. The lack of alignment of prophage remnants between these closely related strains suggests that the prophage decay process must occur over short evolutionary time periods.

In the O157 strain EDL933, 12 prophage sequences were identified. Only the Shiga-toxin-converting phage 933W can produce infectious particles (40). Among the 18 prophages of the O157 Sakai strain, 11 are λ-like phages that share large regions of high DNA sequence identity among themselves and phage λ (figure 3-4). It is currently not clear whether this intriguing similarity is the result of infection by closely related phages or propagation by a copy- and -paste mechanism within this cell lineage followed by some diversification via modular exchanges. Shared prophage sequences are associated with extensive genome rearrangements in *Xylella fastidiosa*, a Gram-negative plant bacterial pathogen known from different pathovars (44).

The remaining O157 prophages distantly resemble phages P2 and P4 and closely resemble phage Mu. They all contain frameshift mutations and various types of deletions and insertions of IS elements and no infectious phage could be induced (33, 51).

The Shiga toxin Stx2- producing prophages in the two O157 strains are integrated into the same locus *wbrA* and their DNA sequences are nearly identical over 85% of the genome (33, 40). They differ over the genetic switch and the DNA replication region. Both phages carry additional virulence factors (*bor, lom*) and a toxin/antitoxin system used by plasmids to ensure their maintenance. Numerous potential virulence factors are also encoded by further prophages from both O157 strains: Shiga toxin Stx1 (933 V, Sp15), an intestinal colonization factor (933 O), and superoxide dismutase (Sp 4, 10) (37). The potential lysogenic conversion genes were located in a few preferred prophage genome positions: in decreasing frequency downstream of the Q-like and N-like antiterminator and the tail fiber genes (5). This observation suggests different acquisition and transcriptional regulation mechanisms in prophages from Gram-negative and Gram-positive bacteria.

Prophages have apparently played a decisive role in the emergence of O157 as a food pathogen. This is not an isolated case. *Salmonella typhimurium*, another important food pathogen, possesses a variable assortment of prophages that apparently represent a transferable repertoire of pathogenic determinants (24). The importance of prophage genes for the in vivo virulence of *Salmonella* was demonstrated by inactivation studies of selected prophage genes (23). In addition, prophages were an important source of genetic diversity between two closely related *Salmonella enterica* strains showing distinct pathogenic potential (serovars Typhimurium and Typhi) (34).

Outlook

This overview can provide only a short outline of an exciting research area at the interface between phage and bacterial genomics, evolutionary biology, and medical microbiology. From the few systems presented here it becomes clear that the interaction between phages and bacteria is not a simple arms race between a parasite and its host. Phage DNA also contributes to the fitness of the bacterial cell. Phages are apparently used by bacterial cells for the rapid acquisition and ecological testing of nonessential "phage" genes that were themselves probably acquired from other bacterial genomes during rare passage of the phage in a heterologous host. This leads to provocative questions. Are phages major drivers of the evolution of bacterial pathogens? Can the variable clinical potential of some protean bacterial pathogens such as *S. pyogenes*, *E. coli*, or *Salmonella* spp. be interpreted by possession of a specific prophage set? Many bacterial pathogens show a very dynamic epidemiology. Is the replacement of older by newer epidemic strains linked to the acquisition of new prophages or prophage combinations? Prophage genes are apparently not a silent cargo to the bacterial lysogen. Upon changed cultivation conditions, prophage genes frequently represent prominently upregulated genes. This was, for example, observed in *S. pyogenes* during temperature shift experiments (42) that mimicked the transition from mucosa-associated bacteria to bacteria in the bloodstream during acute disease. Other examples are the upregulation of prophage gene transcription in pathogenic *E. coli* strains when laboratory-grown bacteria were compared with bacteria during acute infection in an animal (20). Filamentous prophage belonged to the prominent genes showing changes in gene expression when *Pseudomonas aeruginosa* strains were compared in the transition from planktonic to biofilm growth (50). Finally, some phage genes might "cross-talk" with mammalian genes. For example, some lysogenic conversion genes from *S. pyogenes* prophage genes are only expressed when the bacterial cell comes into contact with target cells of the mammalian host (6). The latter excrete a low molecular weight factor that induces the transcription of the prophage gene. It is increasingly becoming clear that pathogens, commensals, and symbionts represent different positions on a continuum of bacteria–host interactions. An indication is the observation that both gut commensals and pathogenic bacteria are under the selection pressure of the mammalian immune system (30, 32). One might therefore suspect that prophages may also play an important role in the ecological adaptation of bacterial commensals and symbionts.

Note added in proof

An update of the literature can be found in: Canchaya, C., G. Fournous, and H. Brüssow. 2004. The impact of prophages on bacterial chromosomes. Mol. Microbiol. 53: 9–18; Brüssow, H., C. Canchaya, and W.-D. Hardt. 2004. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. Microbiol. Mol. Biol. Rev. 68: 560–602.

References

1. Banks, D. J., S. B. Beres, and J. M. Musser. 2002. The fundamental contribution of phages to GAS evolution, genome diversification and strain emergence. Trends Microbiol. 10:515–521.
2. Barondes, J. J., and J. Beckwith. 1995. *bor* gene of phage λ involved in serum resistance, encodes a widely conserved outer membrane protein. J. Bacteriol. 177:1247–1253.
3. Beres, S., G. L. Sylva, K. D. Barbian, B. Lei, J. S. Hoff, et al. 2002. Genome sequence of a serotype M3 strain of group A *Streptococcus*: phage-encoded toxins, the high-virulence phenotype, and clone emergence. Proc. Natl. Acad. Sci. USA 99:10078–10083.
4. Boyce, J. D., B. E. Davidson, and A. J. Hillier. 1995. Identification of prophage genes expressed in lysogens of the *Lactococcus lactis* bacteriophage BK5-T. Appl. Environ. Microbiol. 61:4099–4104.
5. Boyd, E. F., and H. Brüssow. 2002. Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved. Trends Microbiol. 10:521–529.
6. Broudy, T. B., V. Pancholi, and V. A. Fischetti. 2001. Induction of lysogenic bacteriophage and phage-associated toxin from group A streptococci during coculture with human pharyngeal cells. Infect. Immun. 69:1440–1443.

7. Brüssow, H. 2001. Phages of dairy bacteria. Annu. Rev. Microbiol. 55:283–303.
8. Brüssow, H., and F. Desiere. 2001. Comparative phage genomics and the evolution of *Siphoviridae*: insights from dairy phages. Mol. Microbiol. 39:213–222.
9. Brüssow, H., and R. W. Hendrix. 2002. Phage genomics: small is beautiful. Cell 108:113–116.
10. Bruttin, A., F. Desiere, S. Lucchini, S. Foley, and H. Brüssow. 1997. Characterization of the lysogeny module from the temperate *Streptococcus thermophilus* bacteriophage Sfi21. Virology 233:136–148.
11. Campbell, A. M. 1996. Cryptic prophages, pp. 2041–2046. *In* Neidhardt, F. C. (ed.) *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology, 2nd edn. ASM Press, Washington, D.C.
12. Canchaya, C., F. Desiere, W. M. McShan, J. J. Ferretti, J. Parkhill, and H. Brüssow. 2002. Genome analysis of an inducible prophage and prophage remnants integrated in the *Streptococcus pyogenes* strain SF370. Virology 302:245–258.
13. Canchaya, C., C. Proux, G. Fournous, A. Bruttin, and H. Brüssow. 2003. Prophage genomics. Microbiol. Mol. Biol. Rev. 67:238–276.
14. Casjens, S. 2003. Prophages and bacterial genomics: what have we learned so far? Mol. Microbiol. 49:277–300.
15. Chopin, A., A. Bolotin, A. Sorokin, S. D. Ehrlich, and M.-C. Chopin. 2001. Analysis of six prophages in *Lactococcus lactis* IL1403: different genetic structure of temperate and virulent phage populations, Nucleic Acids Res. 29:644–651.
16. Desiere, F., S. Lucchini, and H. Brüssow. 1999. Comparative sequence analysis of the DNA packaging, head, and tail morphogenesis modules in the temperate *cos*-site *Streptococcus thermophilus* bacteriophage Sfi21. Virology 260:244–253.
17. Desiere, F., C. Mahanivong, A. J. Hillier, P. S. Chandry, B. E. Davidson, and H. Brüssow. 2001. Comparative genomics of lactococcal phages: insight from the complete genome sequence of *Lactococcus lactis* phage BK5-T. Virology 283:240–252.
18. Desiere, F., W. M. McShan, D. Van Sinderen, J. J. Ferretti, and H. Brüssow. 2001. Comparative genomics reveals close genetic relationships between phages from dairy bacteria and pathogenic streptococci: evolutionary implications for prophage-host interactions. Virology 288: 325–341.
19. Desiere, F., R. D. Pridmore, and H. Brüssow. 2000. Comparative genomics of the late gene cluster from *Lactobacillus* phages. Virology 275:294–305.
20. Dozois, C. M., F. Daigle, and R. Curtiss. 2003. Identification of pathogen-specific and conserved genes expressed in vivo by an avian pathogenic *Escherichia coli* strain. Proc. Natl. Acad. Sci. USA 100:247–252.
21. Edlin, G., L. Lin, and R. Kudrna. 1975. Lamba lysogens of *Escherichia coli* reproduce more rapidly than non-lysogens. Nature 255:735–737.
22. Ferretti, J. J., W. M. McShan, D. Ajdic, D. J. Savic, K. Lyon, et al. 2001. Complete genome sequence of an M1 strain of *Streptococcus pyogenes*. Proc. Natl. Acad. Sci. USA 98:4658–4663.
23. Figueroa-Bossi, N., and L. Bossi. 1999. Inducible prophages contribute to *Salmonella* virulence in mice. Mol. Microbiol. 33:167–176.
24. Figueroa-Bossi, N., S. Uzzau, D. Maloriol, and L. Bossi. 2001. Variable assortment of prophages provides a transferable repertoire of pathogenic determinants in *Salmonella*. Mol. Microbiol. 39:260–271.
25. Fournous, G. 2003. Automatisation de la recherche et de l'analyse de prophages dans les genomes bacteriens. DESS Université Henri Poincaré, Nancy, France.
26. Hayashi, T., K. Makino, M. Onishi, K. Kurokawa, et al. 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res. 8:11–22.
27. Ikebe, T., A. Wada, Y. Inagaki, K. Sugama, R. Suzuki, et al. 2002. Dissemination of the phage-associated novel superantigen gen *speL* in recent invasive and noninvasive M3/T3 *Streptococcus pyogenes* isolates in Japan. Infect. Immun. 70:3227–3233.
28. Inagaki, Y., F. Myouga, H. Kawabata, S. Yamai, and H. Watanabe. 2000. Genomic differences in *Streptococcus pyogenes* serotype M3 between recent isolates associated with toxic shock-like syndrome and past clinical isolates. J. Infect. Dis. 181:975–983.
29. Jin, Q., J. Yuan, Y. Xu, Y. Wang, Y. Shen, et al. 2002. Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity through comparison with genomes of *Escherichia coli* K12 and O157. Nucleic Acids Res. 30:4432–4441.
30. Krinos, C. M., M. J. Coyne, K. G. Weinacht, A. O. Tzianabos, D. L. Kasper, and L. E. Comstock. 2001. Extensive surface diversity of a commensal microorganism by multiple DNA inversions. Nature 414:555–558.
31. Lawrence, J. G., R. W. Hendrix, and S. Casjens. 2001. Where are the pseudogenes in bacterial genomes? Trends Microbiol. 9:535–540.
32. Macpherson, A., D. Gatto, E. Sainsbury, G. R. Harriman, H. Hengartner, and R. M. Zinkernagel. 2000. A primitive T cell-independent mechanism of intestinal mucosal IgA responses to commensal bacteria. Science 288:2222–2226.
33. Makino, K., K. Yokoyama, Y. Kubota, C. H. Yutsudo, et al. 1999. Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic *Escherichia coli* O157:H7 derived from the Sakai outbreak. Genes Genet. Syst. 74:227–239.
34. McClelland, M., K. E. Sanderson, J. Spleth, and S. W. Clifton. 2001. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. Nature 413:852–856.
35. Morgan, G. J., G. F. Hatfull, S. Casjens, and R. W. Hendrix. 2002. Bacteriophage Mu genome sequence: analysis and comparison with Mu-like prophages in *Haemophilus*, *Neisseria* and *Deinococcus*. J. Mol. Biol. 317:337–359.
36. Nakagawa, I., K. Kurokawa, A. Yamashita, M. Nakata, Y. Tomiyasu, et al. 2003. Genomic sequence of an M3 strain of *Streptococcus pyogenes* reveals a large-scale genomic rearrangement in invasive strains and new insights into phage evolution. Genome Res. 13:1042–1055.
37. Ohnishi, M., K. Kurokawa, and T. Hayashi. 2001. Diversification of *Escherichia coli* genomes: are bacteriophages the major contributors? Trends Microbiol. 9:481–485.

38. Parkhill, J., G. Dougan, K. D. James, and N. R. Thomson. 2001. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. Nature 413:848–852.
39. Perna, N. T., G. Plunkett, V. Burland, and B. Mau. 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature 409:529–533.
40. Plunkett, G., D. J. Rose, T. J. Durfee, and F. R. Blattner. 1999. Sequence of Shiga toxin 2 phage 933W from *Escherichia coli* O157:H7: Shiga toxin as a phage late-gene product. J. Bacteriol. 181:1767–1778.
41. Smoot, J. C., K. D. Barbian, J. J. van Gompel, L. M. Smoot, M. S. Chaussee, et al. 2002. Genome sequence and comparative microarray analysis of serotype M18 group A *Streptococcus* strains associated with acute rheumatic fever outbreaks. Proc. Natl. Acad. Sci. USA 99:4668–4673.
42. Smoot, L. M., J. C. Smoot, M. R. Graham, G. A. Somerville, D. E. Sturdevant, C. A. Migliacco, G. L. Sylva, and J. M. Musser. 2001. Global differential gene expression in response to growth temperature alteration in group A *Streptococcus*. Proc. Natl. Acad. Sci. USA 98:10416–10421.
43. Tettelin H., V. Masignani, M. J. Cieslewicz, J. A. Eisen, S. A. Peterson, et al. 2002. Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V *Streptococcus agalactiae*. Proc. Natl. Acad. Sci. USA 99:12391–12396.
44. Van Slys, M. A., M. C. de Oliviera, C. B. Monteiro-Vitorello, C. Y. Miyaki, and L. R. Furlan. 2003. Comparative analyses of the complete genome sequences of Pierce's disease and citrus variegated chlorosis strains of *Xylella fastidiosa*. J. Bacteriol. 185:1018–1026.
45. Ventura, M., A. Bruttin, C. Canchaya, and H. Brüssow. 2002. Transcription analysis of *Streptococcus thermophilus* phages in the lysogenic state. Virology 302:21–32.
46. Ventura, M., C. Canchaya, D. Pridmore, B. Berger, and H. Brüssow. 2003. Integration and distribution of *Lactobacillus johnsonii* prophages. J. Bacteriol. 185: 4603–4608.
47. Ventura, M., C. Canchaya, M. Kleerebezem, W. M. de Vos, R. J. Siezen, and H. Brüssow. 2003. The prophage sequences of *Lactobacillus plantarum* strain WCFS1. Virology 316: 245–255.
48. Ventura, M., S. Foley, A. Bruttin, S. Chibani-Chennoufi, C. Canchaya, and H. Brüssow. 2002. Transcription mapping as a tool in phage genomics: the case of the temperate *Streptococcus thermophilus* phage Sfi21. Virology 296:62–76.
49. Wagner, P. L., and M. K. Waldor. 2002. Bacteriophage control of bacterial virulence. Infect. Immun. 70:3985–3993.
50. Whiteley, M., M. G. Bangera, R. E. Bumgarner, M. R. Parsek, G. M. Teitzel, S. Lory, and E. P. Greenberg. 2001. Gene expression in *Pseudomonas aeruginosa* biofilms. Nature 413:860–864.
51. Yokoyama, K., K. Makino, Y. Kubota, M. Watanabe, et al. 2000. Complete nucleotide sequence of the prophage VT1-Sakai carrying the shiga toxin 1 genes of the enterohemorrhagic *Escherichia coli* O157:H7 derived from the Sakai outbreak. Gene 258:127–139.

4

Evolution of Tailed Phages: Insights from Comparative Phage Genomics

HARALD BRÜSSOW AND FRANK DESIERE

Many phage researchers believe that phages are as old as their bacterial hosts. If this hypothesis is true, then we have to postulate elements of vertical evolution for phages. In view of the postulated antiquity of these relationships we might not expect sequence similarities between more distantly related phages. Comparative phage genomics can reveal DNA sequence, protein sequence, or gene map similarities in the absence of sequence similarities between increasingly diverged phages. For even more distant relationships, data from structural biology can be informative. Recent genomics-based ideas on phage evolution are dominated by two interpretations. In one view all double-stranded DNA tailed phage genomes are mosaics with access, by horizontal exchange, to a large common genetic pool, but in which access to the gene pool is not uniform for all phages (22). In this hypothesis horizontal gene transfer dominates over vertical evolution. In fact, it is in some way an updated version of the classical modular theory of phage evolution developed 30 years ago on the basis of heteroduplex mapping with lambdoid coliphages (2, 7). Other investigators studying phages from dairy bacteria observed strong elements of vertical evolution in the structural gene cluster of phages that were not erased by horizontal gene transfer events (5). The two hypotheses are not mutually exclusive since they only set a different balance for horizontal and vertical elements in phage evolution. In the following we will present the basic observations leading to the two hypotheses, search for a synthesis of both concepts, and challenge this unifying view with recent phage sequence data.

Comparative Genomics of *Streptococcus thermophilus* Phages

Six *Streptococcus thermophilus* (St) phages were completely sequenced. As revealed by comparative genomics the population genetics of St phages is relatively simple (5) (figure 4-1). The hypothetical St phage genome can be subdivided into four large segments, each with its own mechanisms for creating diversity. One segment is the late gene cluster extending from the DNA packaging genes to the tail genes. This module is represented by two unrelated configurations: one is characteristic for *cos*-site phages (prototype: phage Sfi21), the other for *pac*-site phages (prototype: phage Sfi11) (27). The two structural gene clusters are not related to each other at the nucleotide or protein sequence level. Both clusters diversify by the accumulation of point mutations (28). Pair-wise comparisons within each cluster revealed on average 10–20% bp differences.

A second segment covers the putative tail fiber, lysis, and lysogeny genes. Diversity is created by insertion, deletion, and replacement of DNA segments and to a lesser degree by point mutations (11). In the lysogeny module, recombination processes apparently underlie the acquisition of different types of superinfection immunity and repressor binding specificity in the genetic switch region. When the lysogeny modules from two temperate St phages were aligned, an alternation of conserved and variable DNA segments was observed (37). Some transition zones were exactly at gene borders. Others were in the middle of genes separating protein domains. Deletions in spontaneous or repressor-selected phage mutants were located conspicuously close to these transition zones. Lytic phages, which dominate the St phage population, are apparently derived from temperate phages by a combination of rearrangement and deletion events in the lysogeny module (27). Recombination apparently also plays a role in the creation of diversity in the putative tail fiber genes. In the phage protein that probably interacts with the phage receptor on the bacterial cell, variable and conserved DNA segments alternated (27). Spontaneous deletions were observed that started and ended in DNA repeats encoding collagen-like protein motifs (11). Phages that differed in host range showed completely unrelated variable regions, while phages with overlapping host ranges shared highly related variable regions. Swapping of

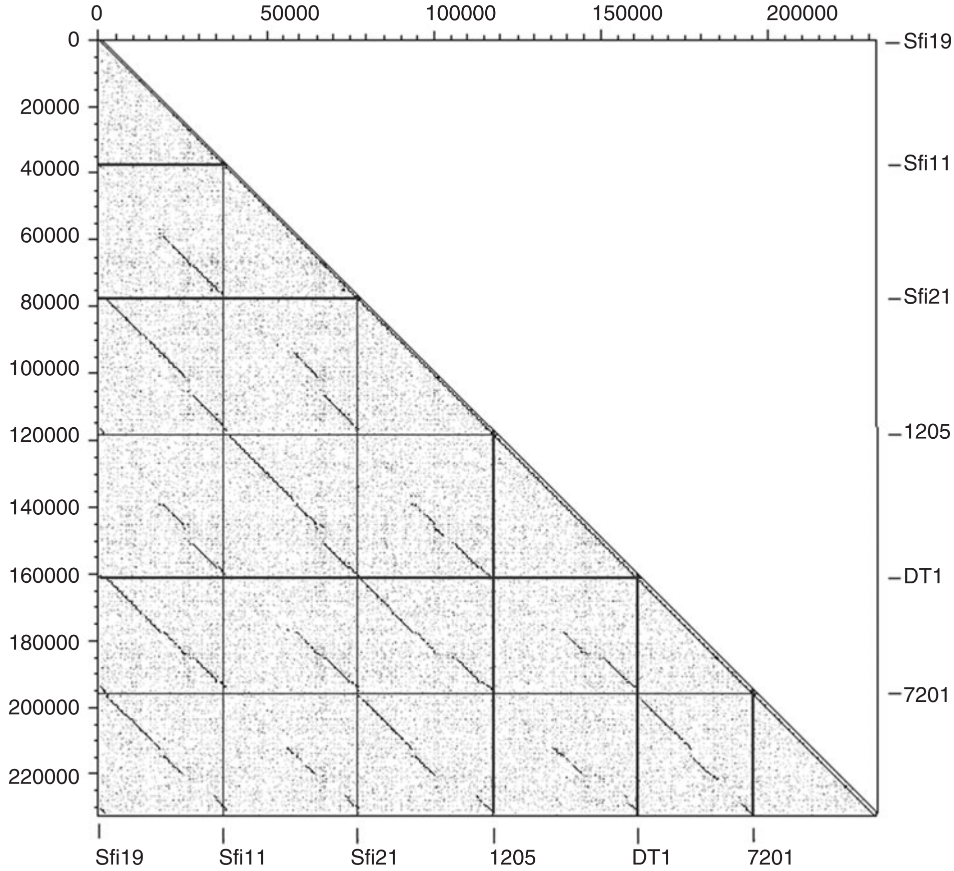


Figure 4-1 Dot-plot matrix calculated for the genome sequences of the *S. thermophilus* phages. Phages include Sfi19, Sfi11, Sfi21, O1205, DT1, and 7201. The left *y*-axis provides a scale in kilobases. The phage sequence on the *x*-axis is identified below the column, that on the *y*-axis to the left of the row. The dot-matrix was calculated using Dotter (45). The comparison window was 50 bp, and the stringency was 30 bp.

variable domains between St phages resulted in corresponding host range changes in the recombinant phage (18).

The third genome segment was the putative DNA replication module represented by two distinct gene constellations: the Sfi21-like and the 7201-like DNA replication module. The Sfi21-like DNA replication module is present in the vast majority of isolated phages and is unusually conserved (13). Even at the third codon position less than 1% sequence diversity was observed in independent phage isolates, suggesting that this module was recently spread horizontally between St phages. The fourth St phage genome segment covers the rightmost 5 kb of the genome. This region gives rise to early transcripts encoding a protein necessary for middle and late transcription (46). Diversity is created by insertion/deletion processes while the DNA sequence is highly conserved (27).

Graded Relatedness in Sfi21-like *Siphoviridae*

Graded relatedness is the hallmark of evolving systems. Comparison of phages infecting the same host species will not yield much evidence for vertical phage evolution. Therefore, comparisons were made between temperate *Siphoviridae* infecting distinct, but evolutionarily related genera of host bacteria. Complete genome sequences are now available for many Sfi21-like *cos*-site *Siphoviridae* that infect several distinct genera of low G-C content, Gram-positive bacteria (14). The comparison of these genomes led to two interesting observations (figure 4-2). First, a gradient of relatedness was detected between these phages. Second, when projected on St phages the relatedness of the phages reflected approximately the phylogenetic relationships of their bacterial hosts, suggesting some coevolution of phages with their host bacteria. The closest relatives of Sfi21 were other *cos*-site *S. thermophilus* phages, which shared more than 80% DNA sequence identity with Sfi21. At the next level of relatedness is the *Lactococcus* phage BK5-T, which shared 60% DNA identity over the DNA packaging and head morphogenesis genes with Sfi21. At the protein level, the similarity between the two phages extended essentially over the entire morphogenesis module (14). The next most closely related phage to Sfi21 is *Lactobacillus* phage adh. Sfi21 and adh are linked by about

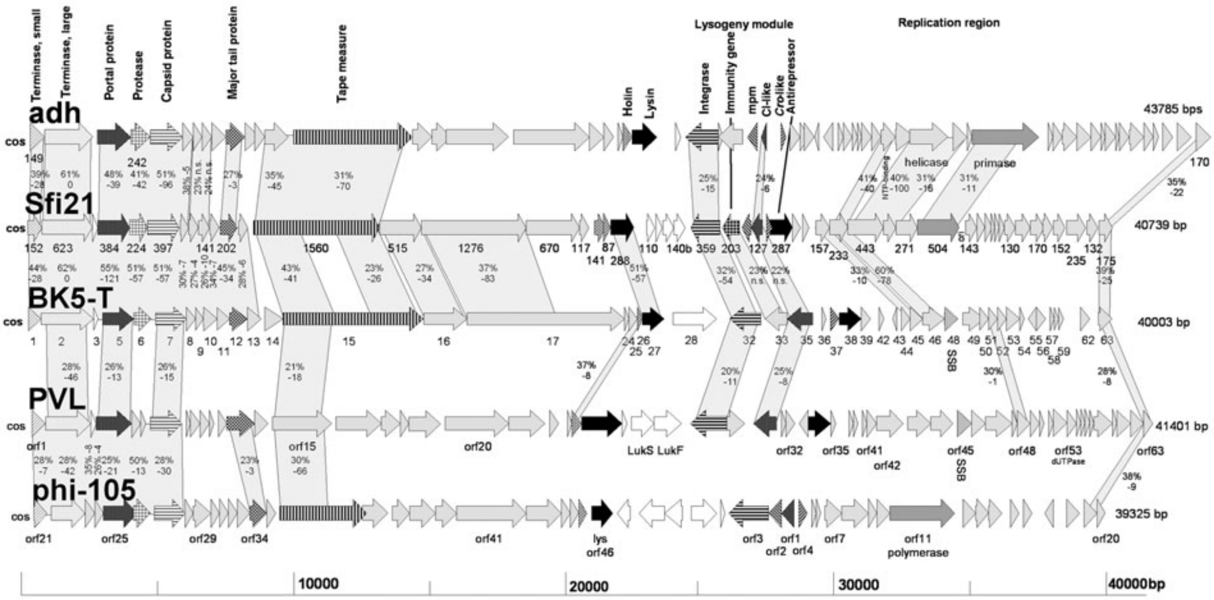

Terminase, small
Terminase, large
Portal protein
Protease
Capsid protein
Major tail protein
Tape measure
Holin
Lysin
Lysogeny module
Integrase
Immunity gene
mpm
CI-like
Cro-like
Antirepressor
Replication region
helicase
primase
NTP-binding
adh
43785 bps
Sfi21
40739 bp
BK5-T
40003 bp
PVL
41401 bp
phi-105
39325 bp
cos
149
242
170
152 623 384 224 397 141 202 1560 515 1276 670 117 87 141 288 110 140b 359 203 127 287 157 233 443 271 504 143 130 170 152 132 235 175
1 2 3 5 6 7 8 9 10 11 12 13 14 15 16 17 24 25 26 27 28 32 33 35 36 37 38 39 42 43 44 45 46 48 49 50 51 52 53 54 55 56 57 58 59 62 63
SSB
orf1 orf15 orf20 LukS LukF orf32 orf35 orf41 orf42 orf45 orf48 orf53 orf63
dUTPase
orf21 orf25 orf29 orf34 orf41 lys orf46 orf3 orf1 orf2 orf4 orf7 orf11 orf20
polymerase
10000
20000
30000
40000bp

Figure 4-2 Alignment of the genetic maps of phages from diverse bacterial lineages. Included are *Lactobacillus* phage adh, *Streptococcus* phage Sfi21, *Lactococcus* phage BK5-T, *Staphylococcus* phage PVL, and *Bacillus* phage ϕ105. Corresponding genes were shaded similarly and were annotated. Genes encoding proteins that showed amino acid sequence similarity are linked by shading, and the percentages of amino acid identity and their log BLAST E-values are given. Selected open reading frames were numbered for an eaiser orientation with the original publications. See the bacteriophages.org/frames_0040.htm for a color version of this figure.

40% amino acid identity over the DNA packaging, head and tail morphogenesis genes. DNA similarity was not detected. Lower levels of protein sequence identity were detected between individual proteins of phage Sfi21 and *Bacillus* phage phi-105 or *Staphylococcus* phage PVL. However, the idea of phage–bacterium co-evolution received a blow when Sfi21-like phages sharing DNA sequence, protein sequence, or only gene map similarity were observed in a single bacterial species (41).

Sfi11-like *Siphoviridae*

Sfi11-like *pac*-site phages were also not limited to *S. thermophilus*. Phages with extensive protein sequence similarity to the structural genes of Sfi11 were detected in a number of *pac*-site phages from low G-C content, Gram-positive bacteria. This series covers, in decreasing order of relatedness, prophages from *S. pyogenes*, and phages from *Lactococcus* (TP901-1, Tuc2009), *Lactobacillus* (phig1e), *Bacillus* (SPP1), and *Listeria* (A118) (15). Nucleotide sequence similarity between *pac*-site St phages infecting the same host species was high and extended over large regions of the genome. DNA sequence similarity was also detected between *S. thermophilus* and *S. pyogenes* phages. However, the degree of similarity was lower and restricted to part of the structural genes. At the protein level, a complex but extensive network linked *pac*-site phage genomes from different species (15). Over the structural gene cluster the Sfi11-like phages showed an almost identical gene map. Only phage SPP1 differed from the other phages by the insertion of supplementary genes at two genome positions. The Sfi11-like phages differed from the Sfi21-like phages by the possession of two major head proteins instead of one, the possession of a scaffold protein and the lack of proteolytic processing of the major head protein.

λ Supergroup of *Siphoviridae*

Genomic similarities are not limited to phages from the lactic acid bacteria and the *Bacillus* branch of Gram-positive bacteria. Over the DNA packaging, head, and tail modules Sfi21-like phages showed clear similarities with *E. coli* phage HK97 (5, 12) or the *Pseudomonas* phage D3 (14). Over this genome region not only did these phages show nearly identical gene maps but they were also linked by sequence similarities over several proteins. The major head proteins of Sfi21 and HK97 lacked any sequence similarity, but showed identical secondary structure predictions and proteolytic processing at identical amino acid positions (12). Interestingly, the gene map of Sfi11-like phages resembled that of phage λ as closely as Sfi21-like phages resembled *E. coli* phage HK97. The significance of these attributions was supported by weak sequence similarities between phage λ and Sfi11-like phages over the major head protein (15). This observation suggests that the Sfi21/HK97-like and Sfi11/ λ-like head modules represent different lineages of

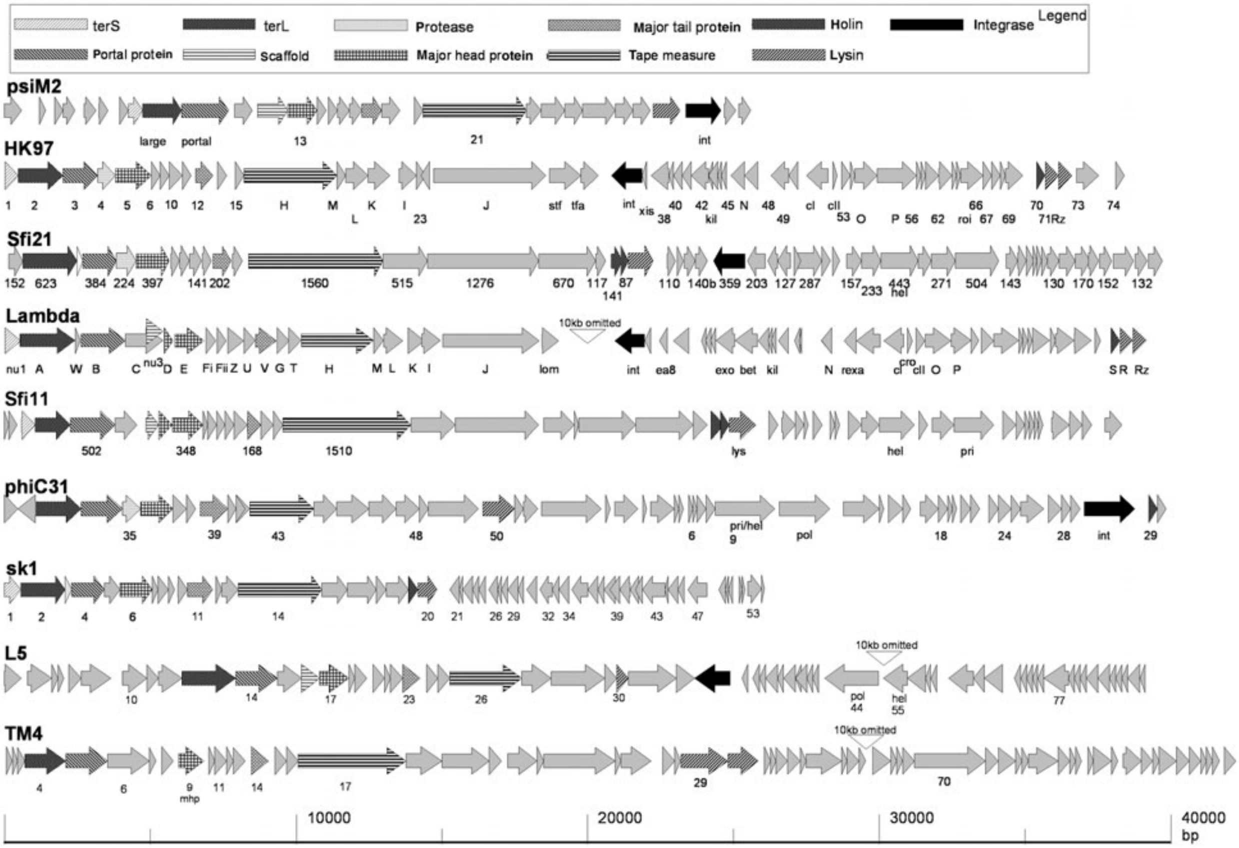

Legend
terS
terL
Protease
Major tail protein
Holin
Integrase
Portal protein
Scaffold
Major head protein
Tape measure
Lysin
psiM2
large portal 13 21 int
HK97
1 2 3 4 5 6 10 12 15 H M K I J stf tfa int xis 40 42 45 N 48 cI cII 66 70 73 74
L 23 38 kil 49 53 O P 56 62 roi 67 69 71Rz
Sfi21
152 623 384 224 397 141 202 1560 515 1276 670 117 87 110 140b 359 203 127 287 157 233 443 271 504 143 130 170 152 132
141 hel
Lambda
10kb omitted
nu1 A W B C nu3 D E FiFii Z U V G T H M L K I J lom int ea8 exo bet kil N rexa cI cro cII O P S R Rz
Sfi11
502 348 168 1510 lys hel pri
phiC31
35 39 43 48 50 6 pri/hel 9 pol 18 24 28 int 29
sk1
1 2 4 6 11 14 20 21 26 29 32 34 39 43 47 53
L5
10kb omitted
10 14 17 23 26 30 pol 44 hel 55 77
TM4
10kb omitted
4 6 9 mhp 11 14 17 29 70
10000 20000 30000 40000 bp

Figure 4-3 Comparison of the genomes from the established and proposed genera of *Siphoviridae* constituting the λ supergroup of phages. Corresponding genes are indicated with the same shading indicated at the top. The following phages were included (from top to bottom): Archaeavirus ψM2, coliphage HK97, *Streptococcus* phage Sfi21, coliphage λ, *Streptococcus* phage Sfi11, *Streptomyces* phage ϕC31, *Lactococcus* phage sk1, Mycobacteria phages L5 and TM4. Selected genes are indicated to allow an easier orientation with the database entries. See the bacteriophages.org/frames_0040.htm for a color version of this figure.

head modules in *Siphoviridae* and have not evolved in either of these two bacterial species. An analysis of the phage genome organization led to the definition of a λ supergroup of *Siphoviridae* (figure 4-3). Notably, the similarities do not cover all *Siphoviridae*: T1-, T5- and SP-like phages are clearly distinct from the λ supergroup. Both observations demonstrate that the mainly morphologically defined taxonomical phage groups do not represent monophyletic groups. Further long-range relationships were detected. For example, Sfi21-like phages differed clearly from lambdoid coliphages in the organization of the lysogeny module, while they shared a number of characteristics with the lysogeny module from P2-like *Myoviridae* from Gram-negative bacteria (29). Surprisingly, tailed phages are found in one branch of the Archaea. With respect to the organization of their structural gene cluster, these viruses are clearly members of the λ supergroup of phages. Notably, archaeaphages showed over their portal and major head proteins clear similarity with a P2-like *Myovirus* (5) and shared related nonstructural genes with a *Lactobacillus* prophage of the Sfi11 series.

Lambdoid Coliphages

The basic ideas on phage evolution were developed for lambdoid coliphages in the pre-genomics era (2, 7). According to the modular theory of phage evolution, the product and unit of phage evolution is not a given virus but a family of interchangeable genetic elements (modules), each of which is multigenic and can be considered as a functional unit. Homologous functions can be fulfilled by a number of distinct DNA segments that lack any sequence similarity. Exchange of a given module for another occurs by recombination among phages belonging to an interbreeding phage population. The theory was developed 30 years ago on the basis of heteroduplex mapping between lambdoid phages (2). This method allowed the distinction of 11 modules in lambdoid coliphages, each represented by a number of alleles. Some modules (e.g., head gene cluster) covered a large DNA segment, were homogeneous, and were represented by only a few distinct alleles. Other modules comprised a small genome region and were represented by many allelic forms (e.g., late gene control) or showed further subdivision into smaller subunits (e.g., early gene control). Comparative genomics of four sequenced lambdoid coliphages (λ, HK022, HK097, N15) identified a nearly perfect colinear gene map for their structural genes (10, 23, 42). Exceptions were small inserts of genes that are under independent transcription control ("morons"). Phages HK097 and HK022 shared extensive DNA sequence identity in a mosaic-like fashion (23). Also phages λ and N15 showed DNA sequence-related structural genes (42). However, the HK097 and λ groups of lambdoid coliphages were not even linked by protein sequence similarity, suggesting that they represent distinct evolutionary lineages of phage structural modules ("Sfi21-/HK097-like" and "Sfi11-/λ-like" lineages) both derived from a very ancient shared ancestor module that has been diversified beyond sequence conservation. Interestingly, *Pseudomonas* phage D3 shared a similar structural gene map and much protein, but no DNA, sequence similarity with HK97 and HK22 (25). We interpret D3 as a more distant relative of HK97 within the Sfi21-like phage lineage. D3 is an interesting evolutionary linker since several of its head proteins were sequence-related to phages from Gram-positive bacteria. In addition, the location of the D3 lysin gene between tail genes and integrase is typical for temperate phages of low G-C content, Gram-positive bacteria.

The right arm of the lambdoid coliphages encoded nonstructural genes. Over this region phages λ, HK022, and HK097 showed a comparable gene organization and shared DNA sequence identity in a patchwise fashion. The fact that phage N15 deviated substantially from this gene map is explained by its peculiar life-style (42): N15 persists in the lysogenic cell as a low-copy linear plasmid with closed hairpin telomeres.

Mechanisms of Recombination

Comparative genomics data can constrain hypotheses on the mechanism of modular exchange reactions. Transition zones from homology to heterology in lambdoid phages were mostly in intergenic regions. As in dairy phages, intragenic transition zones frequently separated protein domains. One report (10) showed the presence of some short regions of sequence homology between distinct nonstructural gene modules in lambdoid coliphages. These linker sequences could promote genetic reassortment (modular exchanges) through homologous or site-specific recombination. In an alternative model nonhomologous recombination occurs indiscriminately and pervasively across the genome of lambdoid phages, followed by stringent selection for functional phages (23). The second part of that process eliminates practically all products of nonhomologous recombination within coding regions, leaving the gene-boundary recombinants and thereby giving the overall process an undeserved appearance of order and purpose. Extensive genome comparisons in mycobacteriophages support this alternative model (40).

Homologous recombination can occur every time a phage infects a cell carrying a prophage with appropriate DNA homologies. These events serve not to create new mosaic boundaries but to rapidly reassort existing gene modules within the population (6). The sequence analysis of two *E. coli* O157 strains underlined the importance of this process. Numerous nearly complete prophage sequences were detected in both O157 strains and many shared long regions of DNA sequence identity over the structural gene cluster with phage λ, one with phage Mu, but none with coliphages HK97, HK620, P2, or P4 (38).

The more than 20 lambdoid coliphage prophage sequences retrieved from the different *E. coli* sequencing projects now allow a first estimate of the genetic diversity within that phage group. For the head gene module, for example, heteroduplex mapping has distinguished four alleles represented by phages λ, 21, ϕ80, and HK97 (7). Three O157 prophages (CP-933N, Sp5, Sp6) lacked DNA sequence similarity with λ or HK97 over the structural module, suggesting new head gene alleles (38). It will be important to determine how many different head gene alleles can be identified in *E. coli*. This figure will in turn determine the theoretical number of permutations that can be achieved by modular exchanges within a given phage group.

Synthesis of Data from Dairy and Lambdoid Phages

Despite the fact that dairy and lambdoid phages infect clearly distinct bacterial hosts (Gram-positive and Gram-negative bacteria), the two phage groups resemble each other in their mosaic relationships, arguing for comparable mechanisms of gene exchange. There are arguments for placing both phage groups in a single supergroup (5). In fact, when restricting the analysis to the structural gene cluster, one could even argue that phage Sfi21 is related more closely to HK97 than the two coliphages HK97 and λ are to each other. The same argument can be made for the relationship between phages Sfi11 and λ. Apparently, modular exchanges between the structural gene clusters from different lambdoid coliphages were not sufficiently intensive to prevent the detection of these relationships. It has been proposed that the intensive protein–protein interaction during virion morphogenesis has impeded over-intensive gene exchanges (8). It is less clear what selective force has maintained the conserved gene order after the postulated split of the ancestral module into the two Sf21/HK97 and Sfi11/λ module lineages.

The nonstructural genes evolved according to other laws. Modular exchanges are much more intensive, resulting in an important reshuffling of genes. It is also clear that distinct structural and nonstructural phage gene clusters are relatively free to combine. For example, nonstructural gene clusters that shared DNA sequence identity were associated in dairy phages with three distinct structural gene clusters (*Lactococcus lactis* phages BK5-T, Tuc2009, and r1t) (4, 14). Similar observations were made in Gram-negative bacteria (phages HK97, λ, P22). In fact, the chimeric origin of phage λ is still apparent from its G-C content, analysis of which showed a clear separation of structural from nonstructural gene clusters. When we speak of phage evolution, we should keep in mind that this term cannot be applied to the entire phage genome (which is a patchwork of different DNA segments with distinct evolutionary histories), but has meaning only for an individual DNA module. However, one can be confident that at least an outline of evolutionary analysis can be made for the structural gene cluster (expressively excluding the tail fiber genes) with a substantially increased phage database. The comparative genomics of the structural gene cluster may also provide a rational basis for a revised phage taxonomy that reflects natural relationships.

Extension to Other Tailed phages?

The phage database is currently still too small to allow definitive conclusions on the mechanisms of diversification and evolution in other phage systems. In particular, claims on a common ancestry for tailed phages (1, 31) or their links to eukaryotic viruses (herpesvirus, insect viruses) are currently not backed by genomics data. However, the available data allow the observation of some trends. First, independent isolates derived from the same or related host species frequently shared DNA sequence identity over essentially the entire genome. Second, this overall alignment was in general interrupted by numerous insertions/deletions and replacements of single genes or small groups of genes. Third, graded relatedness was observed when phages were compared that infected increasingly distant bacterial hosts. In the following, we illustrate these generalizations by examples of phage DNA sequence alignments with the dotter program (figure 4-4).

Siphoviridae

L5- and ψM1-like phages infecting Mycobacteria and Archaea, respectively, are established genera in the *Siphoviridae* phage family. Both showed a more distant, but still detectable similarity in their genome organization with lambdoid phages (5) (figure 4-3). Within the latter group, the two archaeaphages ψM2 and ψM100 differed mainly by one insertion and by sequence diversification over the tail fiber genes (30) (figure 4-4A). In the former group, mycobacteriophages L5 and D29 were well aligned over the entire DNA sequence (19) and differed mainly by a 4 kb-long deletion in D29 (figure 4-4B). The alignment was punctuated by numerous but small insertions/deletions and gene replacements. Larger modular exchanges, in contrast, were not observed. In mycobacteriophages graded relatedness was also observed: phage Bxb1 is a more distant relative of phages L5 and D29 (32). Recent sequencing of 10 further

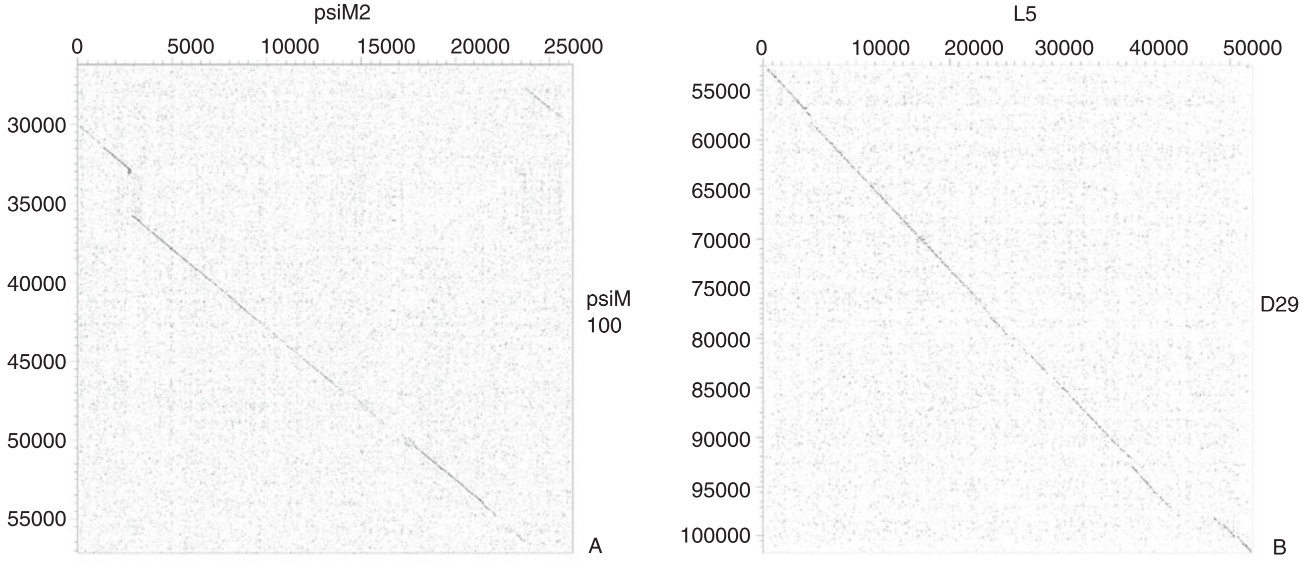


Figure 4-4 Dot-plot analysis of phage couples. Couples include: A: Archaeaphages ψM2 and ψM100; B: Mycobacteriophages L5 and D29; C: *Salmonella* phage P22 and *E. coli* phage HK620; D: *E. coli* phage T7 and *Yersinia* phage ϕYeO3-12; E: *Bacillus* phages GA-1 and PZA; F: *E. coli* phage Mu and prophage Sp18; G: *E. coli* phages P2 and 186; H: *Lactococcus* phages c2 and bIL67. The comparison window was 50 bp pairs and the stringency was 30 bp.

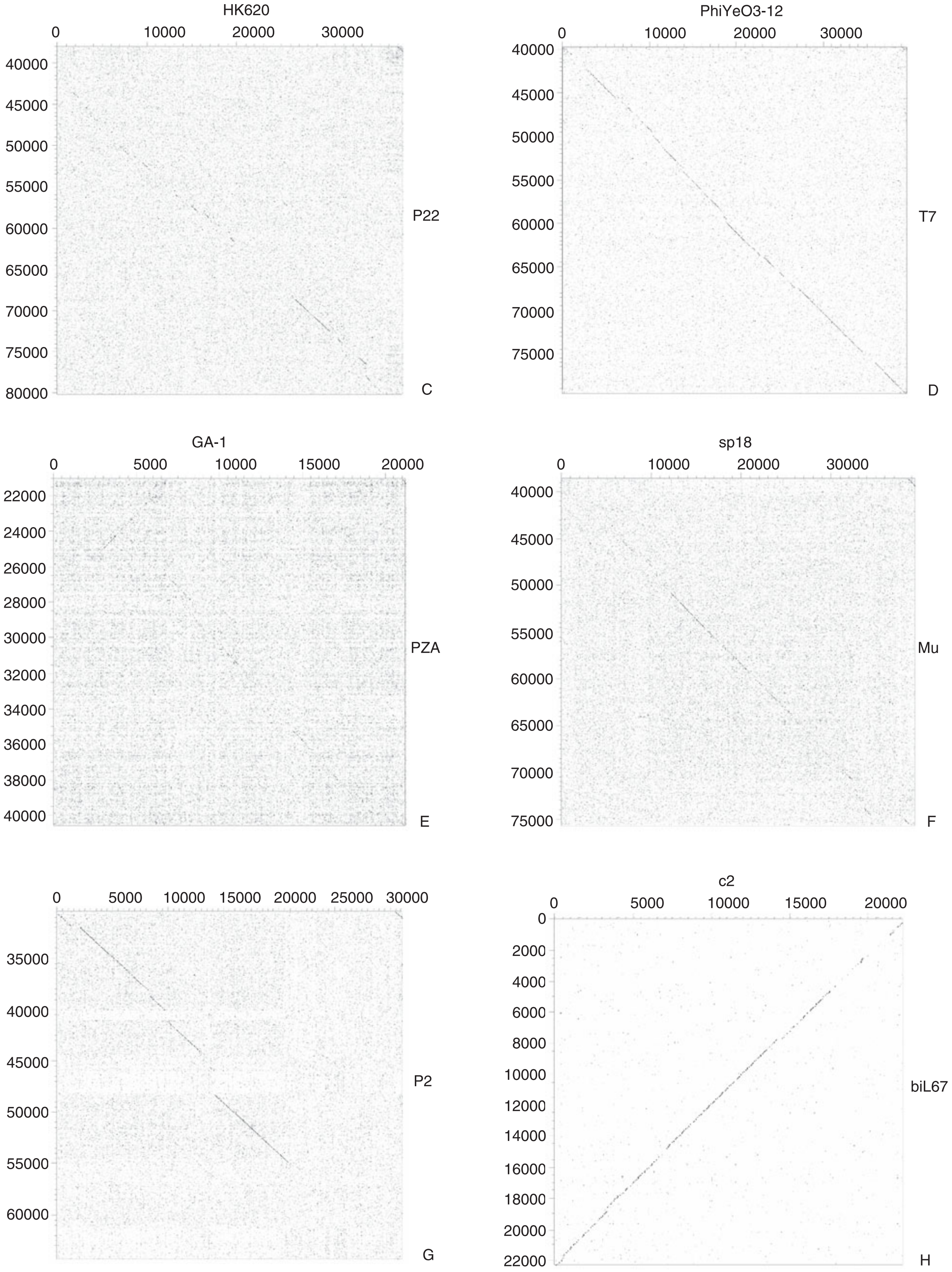


Figure 4-4 Continued.

mycobacteriophages revealed another member of this group, phage Bxz2. However, new groups constituted by related pairs of phage isolates (Che8/Che9d and Omega/Cjw1) or unclassified individual phages were also identified (40).

c2-like *Lactococcus* phages (26), another established genus in *Siphoviridae*, differed substantially from the λ paradigm of structural gene organization, but they could still be aligned with the genome map of the λ-like *Lactococcus* phage sk1 (9) (figure 4-3) when allowing for two genome rearrangement events (4). The two groups of phages still shared sequence relatedness for several predicted proteins (4).

In the DNA alignment *Lactococcus* phages c2 and bIL67 differed mainly over the putative tail fiber region (13, 25) (figure 4-4H). Also, *Lactococcus* phages sk1 and bIL170 could be aligned over the entire DNA sequence and differed mainly by the insertion of several endonuclease genes (see dotter display in 14).

Podoviridae

Taxonomists classify *Salmonella* phage P22 as a podovirus (phages with short, noncontractile tails). P22 demonstrated a graded relationship with *Podoviridae*, infecting increasingly distant hosts. With *E. coli* phage HK620, P22 still shared DNA sequence similarity over large genome regions (figure 4-4C), while with phage APSE-1 infecting a member of the *Proteus* group most of the similarity was limited to the protein sequence level (10, 47, 48). Interestingly, the DNA packaging and head genes from P22-like phages showed a similar genome organization to phage λ; however, protein sequence similarity was no longer detected.

T7-like phages are another genus in the *Podoviridae* family. *E. coli* phage T7 can be aligned with *Yersinia* phage ϕYeO3-12 at the DNA sequence level over essentially the entire genome if one allows for numerous isolated gene replacements (figure 4-4D) (39). In contrast, roseophage SIO1 infecting an evolutionary relative of *Rhodobacter* (α Proteobacteria: *E. coli* and *Yersinia* both belong to the γ-subclass) showed a similar genome organization, and shared protein sequence relatedness but no DNA sequence relatedness with T7 (44). In the ϕ29-like genus of *Podoviridae*, the *Bacillus subtilis* phages GA-1 and PZA could be aligned in a mosaic fashion over most of the genome with one DNA inversion (figure 4-4E). With the *Streptococcus* phage CP-1, PZA shared DNA sequence similarity only over this inversion.

Myoviridae

The same pattern of overall DNA sequence alignment and graded relatedness was also observed in phages with a contractile tail (*Myoviridae*). *E. coli* phage Mu (representative of the Mu-like genus in the *Myoviridae* family) and *E. coli* O157 Sakai prophage Sp18 could be aligned over about half of the genome length at the DNA level in a mosaic fashion (figure 4-4F). Mu shared with the *Shewanella* prophage MuSo2 protein sequence similarity, but only overall genome organization in the absence of sequence similarity when compared with a *Deinococcus* prophage (36).

Due to a larger data set, a more detailed picture can already be painted for the P2-like genus of *Myoviridae*. *E. coli* phages P2 and 186 shared DNA sequence similarity over the entire structural gene cluster except for two extra genes in P2 (one is a lysogenic conversion gene) (17) (figure 4-4G). In comparison, *Haemophilus* phage HP1 shared with P2 sequence-related head proteins but no DNA sequence similarity (16). Vibriophage K139 lacked any sequence similarity with P2 in spite of a conserved genome map.

Also for the T4-like genus of *Myoviridae*, complete phage genome sequences are slowly accumulating. Early heteroduplex analysis of the T-even phages T2, T4, and T6 demonstrated at least 10% DNA sequence divergence between T-even coliphages (24). The heterologous DNA sequences are present as blocks ranging in size from 200 bp to 3 kb. Sequence analysis of selected conserved phage genes such as the capsid gene demonstrated gradients of relatedness leading to the definition of T-evens, Pseudo, Schizo, and Exo T-evens (21, 46). For gp23 this grouping represented protein sequence similarities greater than 90%, 60%, 50%, and 30%, respectively. Different genes led to similar phylogenetic trees, suggesting that in T4-like phages vertical evolution dominated over horizontal evolution, again with the notable exception of the tail fiber genes (20). Complete sequence data are now available for T4 and the T-even phage RB69 and the Pseudo T-even phage RB49 (16, 50). Their genomes are largely colinear with respect to the gene map from T4. Dotplots revealed again a gradient of relatedness. T4 and the T-even phage RB49 could be aligned over nearly the entire genome length at the DNA sequence level. The alignment was, however, interrupted by many small and a few larger gaps (data not shown). One of the larger gaps was around the genes encoding the DNA modifying enzymes of T4. In contrast, T4 and the Pseudo-T-even phage RB49 shared a much lower level of DNA sequence identity and a straight line in their genome alignment was only evident for the structural genes (data not shown). The three T4-like coliphages differ from each other by insertion/deletion or replacements of small groups of genes. At the protein level, RB69 and RB49 had homologues of all essential T4 replication genes, but their sequences diverged considerably from their T4 homologues. In contrast, many of the nonessential T4 genes are absent from RB69 and RB49 and have been replaced by unknown sequences (16, 50). Remarkably, the grouping of the T4-like phages also reflected an increasingly distant relationship of the host bacteria: T-evens were isolated from *Enterobacteriaceae*, Pseudo T-evens covered a larger range of γ Proteobacteria, Schizo T-evens reached into α Proteobacteria, and Exo T-evens, sharing still a low level protein sequence identity with T4, were isolated from Cyanobacteria (21, 46). The Schizo-T-even vibriophage KVP40 shared 36% of the T4 open reading frames, demonstrating 30–60% amino acid identity over the DNA replication, recombination, and repair as well as viral capsid and tail genes. However, its genome is substantially larger than that of T4 (240 vs. 170 kb) and 65% of its open reading frames lacked any database matches (35). The conservation of sequence relatedness in T4-like phage genomes over such large phylogenetic distances of host bacteria is fascinating. λ-like phages infecting γ Proteobacteria and low G-C content, Gram-positive bacteria have lost nearly all sequence similarity and maintained only a related gene map for the structural genes.

Troubles Ahead?

The comprehensive sequencing work of the Pittsburgh Phage Institute revealed substantial sequence diversity and many new types of genome organization in mycobacteriophages (39). Similarly, *Pseudomonas* phage phiKZ, a myovirus with a 280 kb genome, lacked database matches for most of its genes (34). There is thus still substantial *terra incognita* in the sequence space of tailed phages. This observation is also underlined by random sequencing of uncultured virus DNA from seawater: 100 liter samples yielded estimates of up to 7000 different viral sequences (3). Statistical models predicted perhaps a billion different phage-related open reading frames in the oceans (43).

Outlook

We are currently in a transition period. An analysis of 200 complete phage genome sequences in the database has allowed us to perceive some basic principles of phage genome diversification. On a more ambitious scale with several thousand complete phage genomes from a wide evolutionary range of host bacteria, we might arrive at a sequence-based theory of phage evolution and first insights into the relationship of phages to eukaryotic viruses and possible links of phage genomes to the universal tree of life. Alternatively, we might realize that phages are so diverse with respect to their gene content that the data challenge our current taxonomic systems of phage classification and hypotheses on phage evolution. Whatever the outcome, the ease and relatively low cost of phage sequencing, combined with the extensive knowledge on model phages such as λ and T4, could give phage genomics a lead role in population genetics, the evolution of simple DNA genomes, and the modeling of a realistic DNA sequence space.

References

1. Ackermann, H.-W. 1999. Tailed bacteriophages: the order *Caudovirales*. Adv. Virus Res. 51:135–201.
2. Botstein, D. 1980. A theory of modular evolution for bacteriophages, Ann. N.Y. Acad. Sci. 354:484–491.
3. Breitbart, M., P. Salamon, B. Andresen, J. M. Mahaffy, A. M. Segall, D. Mead, F. Azam, and F. Rohwer. 2002. Genomic analysis of uncultured marine viral communities. Proc. Natl. Acad. Sci. USA 99:14250–14255.
4. Brüssow, H. 2001. Phages of dairy bacteria. Annu. Rev. Microbiol. 55:283–303.
5. Brüssow, H., and F. Desiere. 2001. Comparative phage genomics and the evolution of *Siphoviridae*: insights from dairy phages. Mol. Microbiol. 39:213–222.
6. Brüssow, H., and R. W. Hendrix. 2002. Phage genomics: small is beautiful. Cell 108:13–16.
7. Casjens, S., G. F. Hatfull, and R. W. Hendrix. 1992. Evolution of dsDNA tailed-bacteriophage genomes. Semin. Virol. 3:419–429.
8. Casjens, S., and R. W. Hendrix. 1974. Comments on the arrangement of the morphogenetic genes of bacteriophage lambda. J. Mol. Biol. 90:20–23.
9. Chandry, P. S., S. C. Moore, J. D. Boyce, B. E. Davidson, and A. J. Hillier. 1997. Analysis of the DNA sequence, gene expression, origin of replication and modular structure of the *Lactococcus lactis* lytic bacteriophage sk1. Mol. Microbiol. 26:49–64.
10. Clark, A. J., W. Inwood, T. Cloutier, and T. S. Dhillon. 2001. Nucleotide sequence of coliphage HK620 and the evolution of lambdoid phages. J. Mol. Biol. 311:657–679.
11. Desiere, F., S. Lucchini, and H. Brüssow. 1998. Evolution of *Streptococcus thermophilus* bacteriophage genomes by modular exchanges followed by point mutations and small deletions and insertions. Virology 241:345–356.
12. Desiere, F., S. Lucchini, and H. Brüssow. 1999. Comparative sequence analysis of the DNA packaging, head, and tail morphogenesis modules in the temperate *cos*-site *Streptococcus thermophilus* bacteriophage Sfi21. Virology 260:244–253.
13. Desiere, F., S. Lucchini, A. Bruttin, M.-C. Zwahlen, and H. Brüssow. 1997. A highly conserved DNA replication module from *Streptococcus thermophilus* phages is similar in sequence and topology to a module from *Lactococcus lactis* phages. Virology 234:372–382.
14. Desiere, F., C. Mahanivong, A. J. Hillier, P. S. Chandry, B. E. Davidson, and H. Brüssow. 2001. Comparative genomics of lactococcal phages: insight from the complete genome sequence of *Lactococcus lactis* phage BK5-T. Virology 283:240–252.
15. Desiere, F., R. D. Pridmore, and H. Brüssow. 2000. Comparative genomics of the late gene cluster from *Lactobacillus* phages. Virology 275:294–305.
16. Desplats, C., C. Dez, F. Tétart, H. Eleaume, and H. M. Krisch. 2002. Snapshot of the genome of the pseudo-T-even bacteriophage RB49. J. Bacteriol. 184:2789–2804.
17. Dodd, I. B., and J. B. Egan. 1999. P2, 186 and related phages (Myoviridae), pp. 1087–1094. *In* R. Webster, and A. Granoff (eds.) Encyclopedia of Virology. Academic Press, London.

18. Duplessis, M., and S. Moineau. 2001. Identification of a genetic determinant responsible for host specificity in *Streptococcus thermophilus* bacteriophages. Mol. Microbiol. 41:325–336.
19. Ford, M. E., G. J. Sarkis, A. E. Belanger, R.W. Hendrix, and G. F. Hatfull. 1998. Genome structure of mycobacteriophage D29: implications for phage evolution. J. Mol. Biol. 279:143–164.
20. Haggard-Ljungquist, E., C. Halling, and R. Calendar. 1992. DNA sequences of the tail fiber genes of bacteriophage P2: evidence for horizontal transfer of tail fiber genes among unrelated bacteriophages. J. Bacteriol. 174: 1462–1477.
21. Hambly, E., F. Tétart, C. Desplats, W. H. Wilson, H. M. Krisch, and N. H. Mann. 2001. A conserved genetic module that encodes the major virion components in both the coliphage T4 and the marine cyanophage S-PM2. Proc. Natl. Acad. Sci. USA 98:11411–11416.
22. Hendrix, R.W., M. C. Smith, R. N. Burns, M.E. Ford, and G. F. Hatfull. 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. Proc. Natl. Acad. Sci. USA 96:2192–2197.
23. Juhala, R. J., M. E. Ford, R. L. Duda, A. Youlton, G. F. Hatfull, and R. W. Hendrix. 2000. Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. J. Mol. Biol. 299:27–51.
24. Kim, J. S., and N. Davidson. 1974. Electron microscope heteroduplex studies of sequence relations of T2, T4 and T6 bacteriophage DNAs. Virology 57:93–111.
25. Kropinski, A. M. 2000. Sequence of the genome of the temperate, serotype-converting, *Pseudomonas aeruginosa* bacteriophage D3. J. Bacteriol. 182:6066–6074.
26. Lubbers, M.W., N. R. Waterfield, T. P. J. Beresford, R.W. F. Le Page, and A.W. Jarvis. 1995. Sequencing and analysis of the prolate-headed lactococcal bacteriophage c2 genome and identification of the structural genes. Appl. Environ. Microbiol. 61:4348–4356.
27. Lucchini, S., F. Desiere, and H. Brüssow. 1999. Comparative genomics of *Streptococcus thermophilus* phage species supports a modular evolution theory. J. Virol. 73:8647–8656.
28. Lucchini, S., F. Desiere, and H. Brüssow. 1999. The genetic relationship between virulent and temperate *Streptococcus thermophilus* bacteriophages: whole genome comparison of *cos*-site phages Sfi19 and Sfi21. Virology 260:232–243.
29. Lucchini, S., F. Desiere, and H. Brüssow. 1999. Similarly organized lysogeny modules in temperate *Siphoviridae* from low GC content Gram-positive bacteria. Virology 263:427–435.
30. Luo, Y., P. Pfister, T. Leisinger, and A. Wasserfallen. 2001. The genome of archaeal prophage ψM100 encodes the lytic enzyme responsible for autolysis of *Methanothermobacter wolfeii*. J. Bacteriol. 183:5788–5792.
31. Maniloff, J., and H.-W. Ackermann. 1998. Taxonomy of bacterial viruses: establishment of tailed virus genera and the order *Caudovirales*. Arch. Virol. 143:2051–2063.
32. Mediavilla, J., S. Jain, J. Kriakov, M. E. Ford, R. L. Duda, W. R. Jacobs, R.W. Hendrix, and G. F. Hatfull. 2000. Genome organization and characterization of mycobacteriophage Bxb1. Mol. Microbiol. 38:955–970.
33. Meijer, W. J. J., J. A. Horcajadas, and M. Salas. 2001. ϕ29 family of phages. Microbiol. Mol. Biol. Rev. 65:261–287.
34. Mesyanzhinov, V. V., J. Robben, B. Grymonprez, V. A. Kostyuchenko, M. V. Bourkaltseva, N. N. Sykilinda, K. N. Krylov, and G. Volckaert. 2002. The genome of bacteriophage phiKZ of *Pseudomonas aeruginosa*. J. Mol. Biol. 317:1–19.
35. Miller, E. S., J. F. Heidelberg, J. A. Eisen, W. C. Nelson, et al. 2003. Complete genome sequence of the broad-host-range vibriophage KVP40: comparative genomics of a T4-related bacteriophage. J. Bacteriol. 185:5220–5233.
36. Morgan, G.J., G. F. Hatfull, S. Casjens, and R. W. Hendrix. 2002. Bacteriophage Mu genome sequence: analysis and comparison with Mu-like prophages in *Haemophilus*, *Neisseria* and *Deinococcus*. J. Mol. Biol. 317:337–359.
37. Neve, H., K. I. Zenz, F. Desiere, A. Koch, K. J. Heller, and H. Brüssow. 1998. Comparison of the lysogeny modules from the temperate *Streptococcus thermophilus* bacteriophages TP-J34 and Sfi21: implications for the modular theory of phage evolution. Virology 241:61–72.
38. Ohnishi, M., K. Kurokawa, and T. Hayashi. 2001. Diversification of *Escherichia coli* genomes: are bacteriophages the major contributors? Trends Microbiol. 9:481–485.
39. Pajunen, M. I., S. J. Kiljunen, M. E. Soderholm, and M. Skurnik. 2001. Complete genomic sequence of the lytic bacteriophage phiYeO3-12 of *Yersinia enterocolitica* serotype O:3. J. Bacteriol. 183:1928–1937.
40. Pedulla, M. L., M. E. Ford, J. M. Houtz, T. Karthikeyan, C. Wadsworth, J. A. Lewis, N. Jacobs-Sera, et al. 2003. Origins of highly mosaic mycobacteriophage genomes. Cell 113:171–182.
41. Proux, C., D. van Sinderen, J. Suarez, P. Garcia, V. Ladero, G. F. Fitzgerald, F. Desiere, and H. Brüssow. 2002. The dilemma of phage taxonomy illustrated by comparative genomics of Sfi21-like *Siphoviridae* in lactic acid bacteria. J. Bacteriol. 184:6026–6036.
42. Ravin, V., N. Ravin, S. Casjens, M. E. Ford, G. F. Hatfull, and R. W. Hendrix. 2000. Genomic sequence and analysis of the atypical temperate bacteriophage N15. J. Mol. Biol. 299:53–73.
43. Rohwer, F. 2003. Global phage diversity. Cell 113:141.
44. Rohwer, F., A. Segall, G. Steward, V. Seguritan, M. Breittbart, F. Wolven, and F. Azam. 2000. The complete genomic sequence of the marine phage Roseophage SIO1 shares homology with nonmarine phages. Limnol. and Oceanogr. 45:408–418.
45. Sonnhammer, E. L. L., and R. A. Durbin. 1995. Dotmatrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene 167:GC1–10.
46. Tétart, F., C. Desplats, M. Kutateladze, C. Monod, H.-W. Ackermann, and H. M. Krisch. 2001. Phylogeny of the major head and tail genes of the wide-ranging T4-type bacteriophages. J. Bacteriol. 183:358–366.
47. Van der Byl, C., and Kropinski, A. 2000. Sequence of the genome of *Salmonella* bacteriophage P22. J. Bacteriol. 182:6472–6481.
48. Van der Wilk, F., A. M. Dullemans, M. Verbeek, and J. F. van den Heuvel. 1999. Isolation and characterization of APSE-1,

a bacteriophage infecting the secondary endosymbiont of *Acyrthosiphon pisum*. Virology 262:104–113.

49. Ventura, M., S. Foley, A. Bruttin, S. Chibani-Chennoufi, C. Canchaya, and H. Brüssow. 2002. Transcription mapping as a tool in phage genomics: the case of the temperate *Streptococcus thermophilus* phage Sfi21. Virology 296:62–76.
50. Yeh, L.-S., T. Hsu, and J. D. Karam. 1998. Divergence of a DNA replication gene cluster in T4-related bacteriophage RB69. J. Bacteriol. 180:2005–2013.

5

Phage Ecology

STEPHEN T. ABEDON

Phage ecology is the study of the interactions between phage and their environments. These interactions are consequential, particularly to the extent that they affect bacteria. During traditional molecular phage characterization, however, only minimal consideration of ecology is made. Contrasting this tendency, here I consider phage organismal, population, community, and ecosystem ecology (Table 5-1). For additional approaches to the review of phage ecology, as well as the related field of phage environmental microbiology, see (7, 8, 10, 18, 20, 26, 39, 41, 49, 53) together with various reviews of aquatic and ecosystem phage ecology (38, 56, 59, 68, 69, 78). Visit www.phage.org for additional phage ecology resources.

Phage Organismal Ecology

The Basic Phage Life Cycle

While underlain by copious variety and detail, the general phage life cycle involves adsorption, infection, and release, together with a consideration of phage decay (figure 5-1). Adsorption may be further divided into a diffusion-mediated extracellular search, collision between phage and bacterium, attachment between phage and bacterium, and nucleic acid uptake into the bacterial cytoplasm. Infection may be divided into an eclipse period and a period of phage-progeny maturation. The eclipse period is either prevegetative in the sense of immediately proceeding phage-progeny maturation, or is temporarily or greatly extended, as observed, respectively, with pseudolysogeny and lysogeny. Release can occur by various mechanisms, depending on the phage, including via lysis, via extrusion, or via budding. Failure can occur during most or all of these steps, resulting in virion or infection inactivation.

The study of these processes—especially from the perspective of in situ costs, benefits, constraints, expression, and per-infection productivity—is the province of the phage organismal ecologist. More broadly, one can view virus organismal ecology as the study of the adaptations viruses employ to overcome physical, chemical, or biological barriers to transmission between hosts (46). In this section I briefly introduce the phage life cycle as viewed from an ecological perspective.

Phage Adsorption

Phage adsorption begins after phage release from infected cells and ends with the uptake of phage genomes into the cytoplasms of permissive bacterial hosts. The more rapidly phage adsorb, the greater their likelihood of avoiding decay (21, 37, 58) and the shorter their overall life cycle (5). Phage mutants displaying faster adsorption than their parental wild type have been isolated from laboratory cultures (32), which suggests that phage adaptation to new hosts or new conditions can involve evolution of increased adsorption rapidity. This result could also occur if phage adsorption rates in the wild are not always maximal. One probable example of the latter is seen with phage T4 which, by requiring specific adsorption cofactors, may be adsorption-competent within environments in which healthy hosts are likely (e.g., colons) but less adsorption-competent in environments where healthy hosts are less likely (e.g., the extracolonic environment; 31, 49).

Infection (the Latent Period)

The phage latent period begins with the eclipse, during which mature virions are not yet associated with phage-infected bacteria. The post-eclipse phase of an infection can be understood as a period of phage-progeny maturation since mature virions may be released from infected bacteria either following artificial lysis (36) or, as is the case for filamentous phage (see chapter 12), without host lysis. Maturation, for highly virulent phage, mostly occurs at a constant, linear rate rather than exponentially because the rate of synthesis of virion components is limited by the host's anatomy or physiology (42, 74). The situation is complicated, however, if cells are able to continue to grow and divide during phage infection. This is because ongoing growth of a

Table 5-1 Defining Phage Ecology

Ecology	A Bacterium is...	Considerations	Experiments
Organismal	A bacterium is a target, or an entity that impacts on phage phenotype	Phage anatomy, physiology, and behavior characterized from an in situ or a Darwinian perspective; virion stability, survival, and adsorption; eclipse period, latent period, and burst size; adaptations overcoming barriers to transmission between hosts	Single-step growth; adsorption curves; kinetics of phage decay
Population	A bacterium is an environmental resource	Phage population growth and density; liquid versus spatially structured environments (broth growth versus plaque growth); low versus high phage multiplicity; lysogeny versus active phage replication; within-bacteria competition	Batch growth; plaque growth; phage stock preparation
Community	A bacterium is a partner in coevolution	Phage–host coevolution; impact of phage density on density of uninfected bacteria and vice versa; community stability; host resistance; phage host-range breadth and variation; transduction and phage (or lysogenic) conversion; interaction between different phage species	Phage-host continuous-culture or serial-transfer experiments; in situ observation and experiment
Ecosystem	A bacterium is a lower trophic level	Phage impact on ecosystem nutrient cycling and energy flow; short-circuiting of the microbial loop; nutrient release from eukaryote tissues due to attack by phage-encoded toxins	In situ observation and experiment

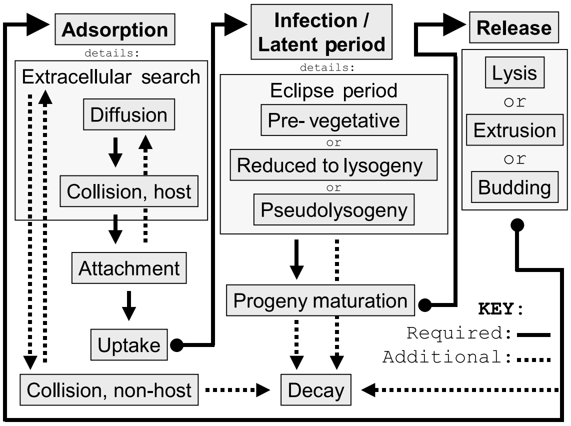


Figure 5-1 General phage life cycle. Continuous lines are paths that must be traversed to complete the phage cycle while dashed lines are optional and may be detrimental to phage propagation. I use the word "Decay" in its most general sense to describe conversion of phage virions or intact phage-infected bacteria from viable plaque-forming units, respectively, to non-infectious virions or to phage-infected bacteria that are unable to produce phage virions.

phage-infected bacterium presumably can replenish those cell components, such as ribosomes, that can limit rates of macromolecule synthesis. At the same time, phage can negatively affect the division of infected bacteria. These negative effects can range from a slowing of host population growth as seen with filamentous phage (48, 63) to a complete cessation of host division as seen with highly virulent phage (e.g., chapter 18).

Phage Progeny Release

Lysis involves a tradeoff between maximizing per-infection phage productivity and minimizing the phage generation time (5, 23). So long as a virus particle remains inside an infected bacterium then it is not free to acquire a new host. For most phage the release of progeny is controlled by phage proteins (e.g., holins) and coincides with the destruction of the parental infected cell (lysis; see chapter 10). For filamentous phage, which extrude their phage progeny across the host cell envelope, release does not necessarily result in host cell death (see chapter 12), thereby allowing filamentous phage to bypass, to some degree, the tradeoff between per-infection productivity and generation time.

Phage Decay

If one is willing to accept that phage are alive (e.g., 27), then phage decay, in its narrowest sense, is equivalent to virion death (or "inactivation" or "loss of titer"; 37). Phage decay (68, 69, 78) likely limits the impact of phage on bacteria (70) and also imposes important constraints on virulent phage: it implies that virion populations cannot survive indefinitely in the absence of sufficient densities of susceptible bacteria (e.g., 30, 76). Similarly, the evolution of lysogeny (see chapter 7) must be dependent at least in part on the relative importance of virion decay versus phage and prophage replication rates as, for example, Stewart and Levin (67) suggest with their "hard times" hypothesis (see also 39). In general, decay impacts directly on phage per-infection productivity by reducing the duration of phage-progeny survival.

Phage Population Ecology

Phage Population Growth

While phage organismal ecology emphasizes virion production and survival (above) and phage community ecology has an emphasis on host population dynamics (below), the emphasis of phage population ecology (table 5-1) is on phage population growth and intraspecific competition, either within bacterial cultures (5) or within individual infected bacteria (28). Like any organism living within a suitable environment possessing sufficient resources, a phage population will increase in number exponentially (2, 24, 34). Phage populations that increase in size most quickly should acquire host cells most rapidly. The acquisition and exploitation of one bacterium by one phage means that one unit of bacterial resource may be less available for exploitation by a second phage. Over the short term, in relatively simple environments, selection within phage populations should therefore be for both more rapid phage population growth and more rapid host cell acquisition.

Certain phage characteristics can contribute to this more rapid growth and host cell acquisition (e.g., 33), though not necessarily without compromise (6, 20). For instance, we should expect evolution to favor decay resistance, particularly so long as this does not require excessive virion sophistication, interfere with rates of phage adsorption, or compete with phage per-infection productivity. Similarly, within more fluid environments we expect selection to favor rapid phage adsorption, though during plaque growth the same expectation is not necessarily realized (as discussed below; see also 32). Since the eclipse period represents a delay until the start of phage-progeny maturation, we also expect selection to result in a shortening of phage eclipse periods. Lysogeny (as well as pseudolysogeny; 12, 55), with its extended delay between infection and progeny maturation, is presumably the product of a countering selection for longer rather than for shorter delays. One also expects a selective favoring of phage that display higher rates of progeny maturation, once the period of progeny maturation has begun, though a maximization of these rates is likely balanced against virion-particle sophistication as well as any host-maintenance costs required for continued infection. Similarly, progeny release, once initiated, should proceed rapidly. Below I consider in greater detail forces that may impact on the evolutionary optimization of the duration of phage latent periods.

Evolution of Phage Latent Period

From an ecological perspective we can divide the members of phage populations into two groups: prereproductive and reproductive. Prereproductive phage are those engaged in either adsorption (including the extracellular search for susceptible bacteria) or the eclipse: during these periods the phage is not generating mature phage progeny. Reproductive phage, by contrast, are those infecting bacteria during the period of phage-progeny maturation. For phage that must lyse their host bacteria to disseminate phage progeny, we may describe a period of progeny maturation as optimal in duration should the latent period giving rise to it result in maximized rates of phage population growth. Latent periods that are too short result in insufficient burst sizes to sustain maximal phage population growth, while those that are too long slow phage population growth by delaying phage-progeny acquisition of uninfected bacteria.

When prereproductive periods are short, this means that free phage can rapidly find uninfected cells and then rapidly gear up for intracellular progeny maturation. Such conditions should select for rapid infection turnover (via shorter latent periods) such that phage progeny acquire uninfected hosts before those cells are obtained by competing phage. In general, then, high host densities and short phage eclipse periods should select for shorter phage latent periods (5). When prereproductive periods are long, by contrast, the reproductive period, once begun, is more valuable, thereby resulting in selection for increased per-infection productivity. Thus, low host densities or long phage eclipse periods should select for larger phage burst sizes, even at the expense of longer phage latent periods (5). The first of these predictions was recently confirmed by competing a mutant of phage RB69 against its longer latent period wild-type parent (6): higher host densities selected for the shorter latent period despite the burst-size cost while lower host densities selected against the mutant. We would expect similar compromises to hold for phage that release their progeny via extrusion, though in those systems the important balance would be between the kinetics of phage release and its impact on both infected-host replication (63) and overall latent-period duration (48). For these latter phage, low densities of phage-susceptible bacteria have been shown to select for higher infected-bacteria fitness, which may be achieved by reducing per-bacterium rates of phage propagation (22). See (23) for a broad discussion of the use of phage latent period and population growth as models for the study of evolutionary optimization from a genetic-mechanism perspective.

Contribution of Early Adsorbers to Phage Population Growth

Phage adsorption occurs essentially as an exponential decay in free-phage density (figure 5-2A). For any phage cohort released at a given moment into a population of potential hosts, phage adsorption occurs such that some constant fraction of remaining free phage will adsorb over any given interval. As a consequence, more phage from a given cohort will adsorb during an earlier interval compared with a later interval. Furthermore, if by chance a phage adsorbs a host earlier rather than later, then the duration for which this phage is prereproductive will be shorter and therefore the total duration of that phage's life cycle will also be shorter. The rate of phage population growth is a function of the duration of the phage life cycle along with the phage burst size. If earlier-adsorbing phage are potentially greater in

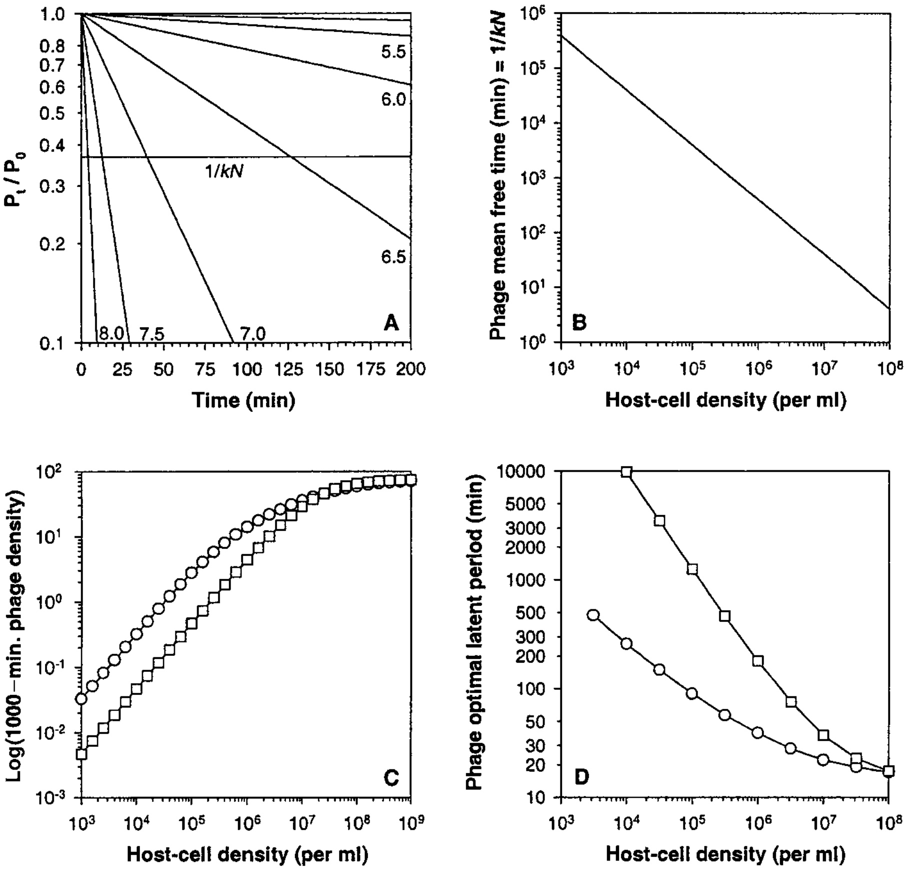


Figure 5-2 Exponential phage adsorption and phage population growth. A: Free-phage adsorption according to the model of Stent (66) with $\log_{10}(N = \text{per ml host density})$ as indicated for different curves and k (the phage adsorption constant) set equal to 2.5×10^{-9} ml/min. Adsorption curves cross the horizontal line at the average phage adsorption time (mean free time) $= 1/kN$. B: Free-phage mean free time graphed as a function of bacterial density. C: Log_{10} phage density following 1000 min of phage growth as simulated or calculated at different host densities (the log scale *and* log transformation of the y-axis are both intentional—the number 10^2 represents a phage density of 10^{100} phage/ml). Assumed are a latent period of 25 min and burst size of 75 phage/cell, together with an adsorption constant 2.5×10^{-9} ml/min. Simulations involve a modeling of *exponential phage adsorption* (circles), which is equivalent to the Stent-modeled adsorption curves in panel A. For calculations it was assumed that individual free phage adsorb after an extracellular search of $1/kN$ min (squares) as calculated as a function of host density in panel B. D: Phage latent periods (optima) that give rise to maximal phage population growth as determined using simulations (meaning adsorption via exponential free-phage decline; circles) or calculations as described for panel C (meaning adsorption for all free-phage cohorts is *mean free time* in duration; squares). Panel D is a variation on a figure originally published in (5) and is used with permission as granted by the American Society for Microbiology. See (5) for discussion of methods for all panels.

number due to the exponential kinetics of phage adsorption, are less susceptible to decay (above), and display shorter life cycles, then it stands to reason that members of phage-virion populations that by chance adsorb earlier will contribute more to phage exponential growth than later-adsorbing members.

At high host densities phage adsorb relatively rapidly such that variance in phage prereproduction duration is not large. However, at lower host densities the timing of the adsorption of the majority of a phage cohort is spread over much longer intervals (figure 5-2A), and the contribution of those phage that by chance are adsorbed by hosts earlier becomes increasingly large and important to overall phage population growth. Thus, the average timing of phage adsorption (the phage mean free time; see 5) declines as a direct function of host density (figure 5-2B), but phage population growth as a function of host density does not decline as quickly (figure 5-2C). As a result, while evolution ought to favor phage with longer replicative periods given lower host densities, optimal latent periods at lower host densities are not nearly as long as would be predicted based on phage mean free times (circles, figure 5-2D; 5).

Phage Plaque Growth

Phage can grow within soft agar overlays—a simple, spatially structured environment—as plaques punctuating otherwise opaque bacterial lawns. Phage growth in plaques may be considered to occur in four stages (47): (i) initial host adsorption of seeded phage, (ii) an initial round of bacterial infection, (iii) an "enlargement phase" that involves multiple rounds of adsorption, infection, and release, and (iv) the end of the enlargement phase, which typically is associated with physiological changes in the bacterial lawn. Differences between phage growth in plaques versus broth occur throughout the enlargement phase, during which the physical structure of solid media (i) slows diffusion, (ii) prevents gross environmental mixing, and (iii) probably gives rise to local phage multiplicities that are higher than those observed over the majority of phage population growth in well-mixed broth media. Phage growth within plaques additionally introduces plaque size as a means by which issues of phage fitness may be addressed (e.g., 51, 52).

We can imagine at least five selective pressures that act on phage during plaque growth: (i) at the periphery of plaques, which is where uninfected bacteria are found, there should be selection for more rapid exponential growth, such as for shorter phage latent periods when host densities are high (above); (ii) regardless of location within a plaque, during the plaque enlargement phase there should be selection for fast diffusion away from the plaque center such that uninfected bacteria surrounding the plaque may be obtained and exploited (similarly, see early interpretations of the classic assertion that smaller virions produce larger plaques than larger virions; e.g., 77); (iii) towards the center of plaques—where there is a low prevalence of uninfected bacteria—there should exist a countering selection for greater burst sizes even at the expense of longer latent periods; (iv) throughout the plaque there should be selection exerted by the tendency of phage to decay (50), including by adsorption to cell debris (61) or adsorption to infected cells (the latter due to superinfection exclusion; 3); and (v) there can be selection for maintenance of phage growth despite the physiological aging of the bacterial lawn (e.g., phage T7; 52). Given this myriad complexity, *how*, *where*, and *when* one determines phage fitness during plaque growth becomes extremely important since different plaque regions may be under different selective pressures. These pressures can vary in addition with host density as well as host physiological state over the course of plaque development.

As a further complication, plaque size does not necessarily correlate with per-infection productivity. It has been hypothesized, for instance, that phage displaying shorter latent periods, even given smaller burst sizes, could display larger plaques (47, 79). Longer latent periods resulting in smaller plaque sizes are most commonly (and classically) observed among T-even phage where lysis-inhibition defective (*r*) mutants display larger plaques and conditionally shorter latent periods than lysis-inhibition competent, wild-type phage (35, 45). Recently phage RB69 mutants have been characterized that display larger plaques than wild-type RB69 along with shorter latent periods and smaller burst sizes (6).

It also has been hypothesized (79) that reducing the likelihood of host attachment given phage–host collision—which, along with rates of phage diffusion, is a major component of the phage adsorption constant (66)—can increase rates of plaque enlargement. The assumption is that with lower attachment efficiency phage spend less time infecting cells and more time diffusing outward from the periphery of plaques. Indeed, one explanation for why phage λ lost its tail fiber upon domestication (44) is that reduced attachment efficiency resulted in the formation of inescapably selectable larger plaques. Sarma and Kaur (64) observed perhaps similar results with various host-range mutants of cyanophage N-1.

Phage Community Ecology

Community Stability

Phage community ecology (table 5-1) emphasizes the bacterial host, for example the impact of phage on bacterial densities and the evolution of phage resistance (9, 16). Phage community ecology also considers phage–host coevolution such as the propensity of phage to evolve strategies that counter mechanisms of host resistance. Bacteria that are resistant to phage attack can contribute to the stability of phage-containing communities by impeding bacterial extinction. Stability additionally refers to the range in densities of host and phage populations as they oscillate over time, with greater oscillation amplitude (density variance) corresponding to lower community stability. Note that reduced community stability—i.e., greater bacterial population instability potentially resulting in local bacterial extinction—is the goal of phage therapy and other means of phage-based biocontrol of bacterial densities (see chapter 48 and reference 40).

Phage community stability in the laboratory is typically studied within the continuous phage–bacterial cocultures commonly referred to as chemostats (60). A chemostat possesses a reservoir containing sterile media that flows to a well-mixed growth vessel containing microorganisms. Flow from reservoir to growth vessel may be controlled via the use of a peristaltic pump, with outflow from the growth vessels occurring at the same rate as inflow. Figure 5-3A presents a simulation of a relatively unstable chemostat. Note that phage have driven phage-sensitive bacteria to extinction after about 110 hours of chemostat progression. Some 100 hours later phage densities decline to extinction due to virion outflow from the chemostat growth

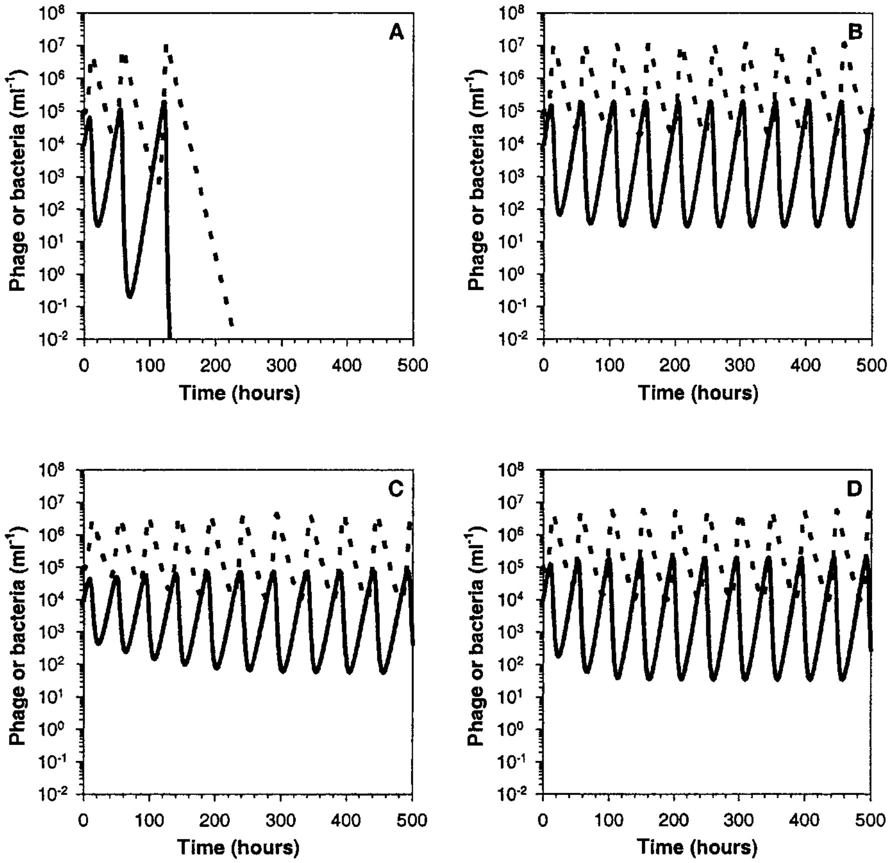


Figure 5-3 Computer-simulated *chemostats*. A: Chemostats were simulated employing the method and parameter values of Bohannan and Lenski (14). Time steps here are 1 min rather than 3 min; the initial host (continuous lines) and phage (dotted lines) densities are 10^4/ml and 10^5/ml, respectively. Unless otherwise noted, the limiting nutrient is glucose that is found in the chemostat reservoir at a density of 0.5 mg/l. B: Shown is a chemostat equivalent to that in panel A though with the phage adsorption constant reduced by one half. C: Shown is a chemostat equivalent to that in panel A but with one half as much limiting glucose present within the virtual chemostat reservoir. D: Shown is a chemostat equivalent to that in panel A except that the phage burst size is reduced by one half. Phage and bacterial densities during simulations were sampled for inclusion in graphs once every 30 min. These simulated chemostats contain no phage-resistant bacteria or other bacterial refuges from phage attack. Extinction is assumed to occur at or below densities of 10^{-2} phage or bacteria/ml.

chamber. Phage–host communities within chemostat, however, are often more stable than may be accounted for by phage community ecology theory (16, 65). Below I consider some of the reasons why.

Refuges

Levin et al. (54) speculated that refuges, which shield sensitive bacteria from phage attack, could increase the stability of phage–host communities, as subsequent experiments have corroborated (65). In such a scheme the phage-induced extinction of sensitive bacteria is prevented by their hiding, for example, within chemostat wall populations. Through cell division these hosts can supply phage-sensitive bacteria to the liquid (unrefuged) phase of the chemostat. Once phage populations have declined due to their outflow from the chemostat, the liquid-phase host population can then grow back to population densities present prior to phage attack.

Slowed Adsorption

Bohannan and Lenski (15) describe bacteria that have entered a "genetic" refuge (phage-resistant mutants) as "invulnerable prey." However, unless a phage's collision with a bacterium results in some degree of phage–host attachment, then a resistant bacterium is not potential prey but instead some relatively inert component of the environment from which phage "bounce." Wilkinson (75), on the other hand, has suggested a model in which completely resistant bacteria really are invulnerable "prey." Here the assumption is that the "predator" species (in this case a *Bdellovibrio* rather than a phage) may reversibly interact with non-prey bacteria by pausing following collision. This delay in detachment extends the *Bdellovibrio*'s extracellular search. From the perspective of susceptible bacteria, this delay is equivalent to a reduction in effective predator density. Wilkinson's conclusion upon modeling such a system is that the presence of invulnerable-prey bacteria, even in the absence of metabolic competition with vulnerable-prey bacteria, will result in an increase in the stability of *Bdellovibrio*-sensitive bacterial densities.

Reductions in phage adsorption rates could similarly result in increased community stability. For instance, a partial reduction in host reception to phage adsorption (e.g., bacteria partially resistant to phage T2) should contribute to an increase in community stability by delaying phage attachment to sensitive bacteria (17). Figure 5-3B presents a simulated chemostat for which the phage adsorption constant has been reduced by one half, and bacteria and phage extinctions are thereby avoided. Again with phage T2, there apparently is a tendency for these phage to be temporarily adsorption-inhibited (up to weeks at room temperature) following release from infected bacteria (62). This phenomenon could also serve to increase community stability by delaying phage adsorption. By reducing phage numbers, mechanisms of phage decay, including outflow from chemostat growth chambers, should also have the effect of increasing community stability. Furthermore, phage evolution could result in a *decrease* in community stability by increasing rates of phage adsorption or by reducing rates of phage decay.

Reduced Phage Productivity

Host density impacts on community stability by affecting the peak phage densities that follow community-wide host lysis. With more phage than hosts within a batch-culture system, eventually all sensitive bacteria may become adsorbed and lysed (10, 29). However, in continuous culture there will be decay in free-phage densities due to outflow from the growth vessel. Since the rate at which host cells are found by free phage is a function of free-phage density (1), there is a race between the survival of phage-sensitive bacteria and free-phage outflow. The lower the peak phage density, the less the bacterial population will be reduced in size due to phage adsorption, and the greater the likelihood that phage adsorption will not reduce the bacterial population to the point of extinction. The smaller the density of bacteria available for infection within a chemostat, in turn, the lower the peak phage density. Bohannan and Lenski (14) demonstrated this point experimentally by reducing the bacterial growth potential through restrictions in the density of a limiting nutrient (glucose) and then observing an increase in phage–bacterial community stability. See figure 5-3C for a simulated chemostat in which the nutrient density in the reservoir has been reduced by one half, and again note that extinction of bacteria and phage is avoided.

The productivity of page infections and infection density together determine peak phage densities. It is well known that phage growth parameters such as lysis timing and burst size can vary as a function of temperature, host nutrition status, phage multiplicity, and host physiology vis-à-vis the standard bacterial growth curve (1–4, 8, 34, 42, 57, 65, 69, 74). If the stability of a chemostat is an inverse function of peak phage density (i.e., more phage = less stability), then reduced infection productivity given reduced nutrient availability should contribute to an increase in community stability. Similarly, we might expect that the T-even-phage lysis-inhibition phenotype (1–4) would be destabilizing since it contributes, particularly at higher host densities, to a larger phage burst size. In figure 5-3D the impact on community stability of reducing the phage burst size by one half is explored, with bacteria and phage extinction yet again avoided.

Synthesis

It is highly likely that phage–host community stability arises from two relatively simple forces: (i) If bacteria cannot be driven to extinction by even excess phage densities, for example as is at least approximated with host refuges from phage attack, then bacteria simply will not be driven to extinction by phage. (ii) If sensitive hosts can be driven to extinction given sufficient phage densities, then hosts will be driven to extinction only if sufficient phage densities are present within an ecosystem. There are two corollaries to the second point: (a) At peak phage density, the fewer phage found within an ecosystem, the smaller the negative impact those phage will have on phage-susceptible bacterial populations and therefore the more stable the system (compare figure 5-3A with figures 5-3C and 5-3D). (b) Mechanisms that interfere with a phage's attainment of higher peak densities (or with the phage impact on individual bacteria)—such as greater phage decay, more rapid outflow, partial inhibition of phage adsorption, or reduced phage burst sizes—can lead to an increase in the stability of a phage–bacterial community (figure 5-3). Indeed, as noted by E. S. Anderson in 1957 (p. 205; 10), "Anything which restricts the phage titre limits the selective action of phage."

Phage Ecosystem Ecology

Phage ecosystem ecology (table 5-1) encompasses the biotic as well as the abiotic world, in particular the biogeochemical cycling of nutrients and the flow of energy between and through ecosystems, usually with an aquatic emphasis (for recent reviews see: 38, 56, 59, 68, 69, 78). Bacteria consume, produce, and store nutrients and energy, as well as contribute to the decomposition of other organisms. Phage infections contribute to a solubilization of bacteria, whether following host-cell lysis or via the conversion of host components into virion particles. Solubilized bacteria, in addition to no longer functioning as consumers, producers, or decomposers, are also less available as food to bacteria grazers: those protists or animals that obtain their nutrients and energy through the ingestion or engulfment of intact bacteria. Since bacteria are the chief consumers of the solubilized components of aquatic ecosystems, a major consequence of the phage-induced lysis of bacteria is not just a reduction in the productivity of bacterial populations but also a delay in the movement up food chains of bacteria-contained nutrients and energy. One can additionally consider the phage-encoding of bacterial toxins (see chapter 47) from the perspective of ecosystem ecology: toxin exposure can free up nutrients from the tissues of bacteria-infected eukaryotes (see 20 for a review of bacterial pathology from the perspective of phage ecology).

In their requirement for intact bacteria, phage in a sense are competitors of the bacteria grazers. Due to the host-range constraints observed among all parasites, individual phage also tend to be more specialized than most grazers in terms of which bacteria within a community they may affect (e.g., 19, 71). The greater the densities of phage-susceptible bacteria—except to the point where bacterial physiology suffers—the faster phage populations can grow. The more phage ultimately produced, the lower the density of phage-susceptible bacteria after a phage attack. Presence of phage within an ecosystem, particularly phage displaying relatively narrow host ranges, consequently has the perverse effect of selecting *against* common bacterial phenotypes, and thereby *for* increased diversity in the bacterial community; that is, for multiple phage-susceptibility types (72, 78). Phage additionally can make direct, positive contributions to the fitness of bacteria hosts through phage conversion (12) or via the transduction of genes from other bacteria (20). Phage DNA and protein coats, following failed infection, might even serve as bacterial nutrients (38).

Phage ecosystem ecology represents an elaboration on the various issues of organismal, population, and community ecology already considered. It follows, therefore, that many or all of the complications discussed throughout this chapter also affect our understanding of the phage impact on ecosystem nutrient cycling and energy flow. In addition, much of the impact of phage on ecosystems, though still incompletely characterized, has been discerned from the study of aquatic phage biology. The influence of phage, especially at microscales, is less easily grasped in more complex ecosystems such as undisturbed soils (25) or in biofilms (43, 73). Indeed, one essential step in characterizing the phage impact on soil bacteria—the determination of a soil-phage total count (e.g., 13)—has only recently been accomplished (11). Thus, both literally and figuratively our understanding of the impact of phage on real ecosystems has barely scratched the surface of phage ecosystem ecology's ultimate goal: quantifying the phage impact on nutrient cycling and energy flow throughout the biosphere.

Acknowledgments

I thank Hans Ackermann and Cameron Thomas for their suggestions and Paul Hyman for lending an ear (and a lab) during the writing process.

References

1. Abedon, S. T. 1990. Selection for lysis inhibition in bacteriophage. J. Theor. Biol. 146:501–511.
2. Abedon, S. T. 1992. Lysis of lysis-inhibited bacteriophage T4 infected cells. J. Bacteriol. 174:8073–8080.
3. Abedon, S. T. 1994. Lysis and the interaction between free phages and infected cells, pp. 397–405. *In* J. D. Karam (ed.) The Molecular Biology of Bacteriophage T4. ASM Press, Washington, D.C.
4. Abedon, S. T. 1999. Bacteriophage T4 resistance to lysis-inhibition collapse. Genet. Res. 74:1–11.
5. Abedon, S. T., T. D. Herschler, and D. Stopar. 2001. Bacteriophage latent-period evolution as a response to resource availability. Appl. Environ. Microbiol. 67: 4233–4241.
6. Abedon, S. T., P. Hyman, and C. Thomas. 2003. Experimental examination of bacteriophage latent-period evolution as a response to bacterial availability. Appl. Environ. Microbiol. 69:7499–7506.
7. Ackermann, H.-W. 1997. Bacteriophage ecology, pp. 335–339. *In* M. T. Martins, M.1. Z. Sato, J. M. Tiedje, L. C. N. Hagler, J. Dobereiner, and P. S. Sanchez (eds.) Progress in Microbial Ecology (Proceedings of Seventh International Symposium on Microbial Ecology). Brazilian Society for Microbiology.
8. Ackermann, H.-W., and M. S. DuBow. 1987. Viruses of Prokaryotes. CRC Press, Boca Raton, Fla.
9. Allison, G. E., and T. R. Klaenhammer. 1998. Phage resistance mechanisms in lactic acid bacteria. Int. Dairy J. 8:207–226.
10. Anderson, E. S. 1957. The relations of bacteriophages to bacterial ecology, pp. 189–217. *In* R. E. O. Williams and C. C. Spicer (eds.) Microbial Ecology. Cambridge University Press, London.
11. Ashelford, K. E., M. J. Day, and J. C. Fry. 2003. Elevated abundance of bacteriophage infecting bacteria in soil. Appl. Environ. Microbiol. 69:285–289.
12. Barksdale, L., and S. B. Ardon. 1974. Persisting bacteriophage infections, lysogeny, and phage conversions. Annu. Rev. Microbial. 28:265–299.
13. Bergh, Ø., K. Y. Børsheim, G. Bratbak, and M. Heldal. 1989. High abundance of viruses found in aquatic environments. Nature 340:467–468.
14. Bohannan, B. J. M., and R. E. Lenski. 1997. Effect of resource enrichment on a chemostat community of bacteria and bacteriophage. Ecology 78:2303–2315.
15. Bohannan, B. J. M., and R. E. Lenski. 1999. Effect of prey heterogeneity on the response of a food chain to resource enrichment. Am. Nat. 153:73–82.
16. Bohannan, B. J. M., and R. E. Lenski. 2000. Linking genetic change to community evolution: insights from studies of bacteria and bacteriophage. Ecol. Lett. 3:362–377.

17. Bohannan, B. J. M., and R. E. Lenski. 2000. The relative importance of competition and predation varies with productivity in a model community. Am. Nat. 156:329–340.
18. Boucher, I., and S. Moineau. 2001. Phages of *Lactococcus lactis*: an ecological and economical equilibrium. Rec. Res. Dev. Virol. 3:243–256.
19. Bratbak, G., T. F. Thingstad, and M. Heldal. 1994. Viruses and the microbial loop. Microb. Ecol. 28:209–221.
20. Breitbart, M., F. Rohwer, and S. T. Abedon. 2004. Phage ecology and bacterial pathogenesis. *In* M. Waldor, D. Friedman, and S. Adhya (eds.) Phages: Their role in Pathogenesis and Biotechnology. ASM Press, Washington, D.C.
21. Bremermann, H.-J. 1983. Parasites at the origin of life. J. Math. Biol. 16:165–180.
22. Bull, J. J., and I. J. Molineux. 1992. Molecular genetics of adaptation in an experimental model of cooperation. Evolution 46:882–895.
23. Bull, J. J., D. W. Pfening, and I.-W. Wang. 2004. Genetic details, optimization and phage life histories. Trends Ecol. Evo1. 19:76–82.
24. Burnet, F. M. 1934. The bacteriophages. Biol. Rev. Cambridge Phil. Soc. 9:332–350.
25. Burroughs, N. J., P. Marsh, and E. M. H. Wellington. 2000. Mathematical analysis of growth and interaction dynamics of streptomycetes and a bacteriophage in soil. Appl. Environ. Microbiol. 66:3868–3877.
26. Campbell, A. 1961. Conditions for the existence of bacteriophages. Evolution 15:153–165.
27. Chao, L. 2000. The meaning of life. BioScience 50:245–250.
28. Chao, L., K. A. Hanley, C. L. Burch, C. Dahlberg, and P. E. Turner. 2000. Kin selection and parasite evolution: higher and lower virulence with hard and soft selection. Q. Rev. Biol. 75:261–275.
29. Chao, L., B. R. Levin, and F. M. Stewart. 1977. A complex community in a simple habitat: an experimental study with bacteria and phage. Ecology 58:369–378.
30. Chopin, A., A. Bolotin, A. Sorokin, S. D. Ehrlich, and M. C. Chopin. 2001. Analysis of six prophages in *Lactococcus lactis* IL1403: different genetic structure of temperate and virulent phage populations. Nucleic Acids Res. 29: 644–651.
31. Conley, M. P., and W. B. Wood. 1975. Bacteriophage T4 whiskers: a rudimentary environment-sensing device. Proc. Natl. Acad. Sci. USA 72:3701–3705.
32. Daniels, L. L., Viand A. C. Wais. 1998. Virulence of phage populations infecting *Halobacterium cutirubrum*. FEMS Microbiol. Ecol. 25:129–134.
33. DeFilippis, V. R., and L. P. Villarreal. 2000. An introduction to the evolutionary ecology of viruses, pp. 125–208. *In* C. J. Hurst (ed.) Viral Ecology. Academic Press, San Diego.
34. Delbřuck, M. 1942. Bacterial viruses (bacteriophages). Adv. Enzymol. 2:1–32.
35. Doermann, A. H. 1948. Lysis and lysis inhibition with *Escherichia coli* bacteriophage. J. Bacteriol. 55:257–275.
36. Doermann, A. H. 1952. The intracellular growth of bacteriophages. I. Liberation of intracellular bacteriophage T4 by premature lysis with another phage or with cyanide. J. Gen. Physiol. 35:645–656.
37. Douglas, J. 1975. Bacteriophages, pp. 77–133. Chapman and Hall, London.
38. Fuhrman, J. A. 1999. Marine viruses and their biogeochemical and ecological effects. Nature 399:541–548.
39. Gill, J. J., and S. T. Abedon. 2003. Bacteriophage ecology and plants. APSnet Feature http://www.apsnet.org/online/feature/phages/
40. Goodridge, L., and S. T. Abedon. 2003. Bacteriophage biocontrol and bioprocessing: application of phage therapy to industry. SIM News 53:254–262.
41. Goyal, S. M., C. P. Gerba, and G. Bitton. 1987. Phage Ecology. CRC Press, Boca Raton, Fla.
42. Hadas, H., M. Einav, I. Fishov, and A. Zaritsky. 1997. Bacteriophage T4 development depends on the physiology of its host *Escherichia coli*. Microbiology 143:179–185.
43. Hanlon, G. W., S. P. Denyer, C. J. Olliff, and L. J. Ibrahim. 2001. Reduction in exopolysaccharide viscosity as an aid to bacteriophage penetration through *Pseudomonas aeruginosa* biofilms. Appl. Environ. Microbiol. 67:2746–2753.
44. Hendrix, R. W., and R. L. Duda. 1992. Bacteriophage lambda PaPa: not the mother of all lambda phages. Science 258:1145–1148.
45. Hershey, A. D. 1946. Mutation of bacteriophage with respect to type of plaque. Genetics 31:620–640.
46. Hurst, C. J., and H. D. A. Lindquist. 2000. Defining the ecology of viruses, pp. 3–40. *In* C. J. Hurst (ed.) Viral Ecology. Academic Press, San Diego.
47. Koch, A. L. 1964. The growth of viral plaques during the enlargement phase. J. Theor. Biol. 6:413–431.
48. Kuo, M. Y., M. K. Yang, W. P. Chen, and T. T. Kuo. 2000. High-frequency interconversion of turbid and clear plaque strains of bacteriophage f1 and associated host cell death. Can J. Microbiol. 46:841–847.
49. Kutter, E., E. Kellenberger, K. Carlson, S. Eddy, J. Neltzel, L. Messinnger, J. North, and B. Guttman. 1994. Effects of bacterial growth conditions and physiology on T4 infection, pp. 406–418. *In* J. D. Karam (ed.) The Molecular Biology of Bacteriophage T4. ASM Press, Washington, D.C.
50. Lee, Y., S. D. Eisner, and J. Yin. 1997. Antiserum inhibition of propagating viruses. Biotech. Bioeng. 55:542–546.
51. Lee, Y., and J. Yin. 1996. Detection of evolving viruses. Nat. Biotechnol. 14:491–493.
52. Lee, Y., and J. Yin. 1996. Imaging the propagation of viruses. Biotech. Bioeng. 52:438–442.
53. Lenski, R. E. 1988. Dynamics of interactions between bacteria and virulent bacteriophage. Adv. Microbial. Ecol. 10:1–44.
54. Levin, B. R., F. M. Stewart, and L. Chao. 1977. Resource limited growth, competition, and predation: a model and experimental studies with bacteria and bacteriophage. Am. Nat. 111:3–24.
55. Los, M., G. Wegrzyn, and P. Neubauer. 2003. A role for bacteriophage T4 *rI* gene function in the control of phage development during pseudolysogeny and in slowly growing host cells. Res. Microbiol. 154:547–552.
56. Martin, E. L., and T. A. Kokjohn. 1999. Cyanophages, pp. 324–332. *In* A. Granoff and R. G. Webster (eds.) Encyclopedia of Virology, 2nd edn. Academic Press, San Diego.
57. Middelboe, M. 2000. Bacterial growth rate and marine virus–host dynamics. Microb. Ecol. 40:114–124.
58. Murray, A. G., and G. A. Jackson. 1992. Viral dynamics: a model of the effects of size, shape, motion, and abundance

of single-celled planktonic organisms and other particles. Mar. Ecol. Prog. Ser. 89:103–116.

59. Paul, J. H., and C. A. Kellogg. 2000. Ecology of bacteriophages in nature, pp. 211–246. *In* C. J. Hurst (ed.) Viral Ecology. Academic Press, San Diego.
60. Paynter, M. J. B., and H. R. Bungay III. 1971. Characterization of virulent bacteriophage infections of *Escherichia coli* in continuous culture. Science 172:405.
61. Rabinovitch, A., I. Aviram, and A. Zaritsky. 2003. Bacterial debris—an ecological mechanism for coexistence of bacteria and their viruses. J. Theor. Biol. 224:377–383.
62. Sagik, B. P. 1954. A specific reversible inhibition of bacteriophage T2. J. Bacteriol. 68:430–436.
63. Salivar, W. O., H. Tzagoloff, and D. Pratt. 1964. Some physical-chemical and biological properties of the rod-shaped coliphage M13. Virology 24:359–371.
64. Sarma, T. A., and B. Kaur. 1997. Characterization of host-range mutants of cyanophage N-1. Acta Virol. 41:245–250.
65. Schrag, S., and J. E. Mittler. 1996. Host–parasite persistence: the role of spatial refuges in stabilizing bacteria–phage interactions. Am. Nat. 148:348–347.
66. Stent, G. 1963. Molecular Biology of Bacterial Viruses. W. H. Freeman, San Francisco.
67. Stewart, F. M., and B. R. Levin. 1984. The population biology of bacterial viruses: Why be temperate? Theor. Pop. Biol. 26:93–117.
68. Suttle, C. A. 2000. Cyanophages and their role in the ecology of cyanobacteria, pp. 563–589. *In* B. A. Whitton and M. Potts (eds.) The Ecology of Cyanobacteria: Their Diversity in Time and Space. Kluwer, Boston, Mass.
69. Suttle, C. A. 2000. The ecology, evolutionary and geochemical consequences of viral infection of cyanobacteria and eukaryotic algae, pp. 248–286. *In* C. J. Hurst (ed.) Viral Ecology. Academic Press, San Diego.
70. Suttle, C. A., and F. Chen. 1992. Mechanisms and rates of decay of marine viruses in seawater. Appl. Environ. Microbiol. 58:3721–3729.
71. Thingstad, T. F. 2000. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. Limnol. Oceanogr. 45:1320–1328.
72. Thingstad, T. F. and R. Lignell. 1997. Theoretical models for control of bacterial growth rate, abundance, diversity and carbon demand. Aquat. Microb. Ecol. 13:19–27.
73. Webb, J. S., L. S. Thompson, S. James, T. Charlton, T. Tolker-Nielsen, B. Koch, M. Givskov, and S. Kjelleberg. 2003. Cell death in *Pseudomonas aeruginosa* biofilm development. J. Bacteriol. 185:4585–4592.
74. Webb, V., E. Leduc, and G. B. Spiegelman. 1982. Burst size of bacteriophage SP82 as a function of growth rate of its host *Bacillus subtilis*. Can. J. Microbiol. 28:1277–1280.
75. Wilkinson, M. H. F. 2001. Predation in the presence of decoys: an inhibitory factor on pathogen control by bacteriophages or bdellovibrios in dense and diverse ecosystems. J. Theor. Biol. 208:27–36.
76. Williams, S. T., A. M. Mortimer, and L. Manchester. 1987. Ecology of soil bacteriophages, pp. 157–179. *In* S. M. Goyal, C. P. Gerba, and G. Bitton (eds.) Phage Ecology. Wiley, New York.
77. Wilson, G. S., and A. A. Miles. 1946. The bacteriophage, pp. 325–350. *In* Topley and Wilson's Principles of Bacteriology and Immunity. Williams and Wilkins, Baltimore.
78. Wommack, K. E., and R. R. Colwell. 2000. Virioplankton: viruses in aquatic ecosystems. Mlcrobiol. Mol. Biol. Rev. 64:69–114.
79. Yin, J., and J. S. McCaskill. 1992. Replication of viruses in a growing plaque: a reaction-diffusion model. Biophys. J. 61:1540–1549.

PART II

LIFE OF PHAGES

This page intentionally left blank

6

DNA Packaging in Double-Stranded DNA Phages

PAUL J. JARDINE
DWIGHT L. ANDERSON

Imagine trying to stuff a string more than 6μm long into a sphere that is fifty nanometers in diameter. The hole in the sphere that the string must enter is only twice as wide as the string itself. The string is stiff, with a persistence length on the order of 50 nm. It is also negatively charged and self-repulsive. The string must be organized such that it can be pulled out easily, so no knots or tangles are permitted. When the sphere is full, the string will have a near crystalline density. You have several minutes to complete this task. This difficult feat is the challenge presented to dsDNA phages during DNA packaging, a pivotal event in the assembly cascade.

The task of compacting the double-stranded DNA chromosome into a protein capsid is a dramatic endeavor. DNA by its nature does not want to be in condensed form, but rather is dispersed, occupying a volume more than 100 times its volume inside the virion (47, 54). Therefore, in order to be packaged, energy must be invested in the DNA. The DNA packaging event must also be coordinated with the replication of the phage DNA that is to be packaged, as well as the assembly and maturation of the protein capsid. Numerous investigators, using a battery of model phage systems, have made a concerted effort over four decades to resolve the components and mechanism of DNA packaging.

Descriptions of the specific components and processes involved in DNA packaging for many of the phages are described in the accompanying chapters of this book. Our intention here is to describe the specific challenges of double-stranded DNA packaging in bacteriophages and detail the common events and structures involved. For most of the systems dealt with here, an extensive battery of biochemical and genetic resources has accumulated over the past half century. Defined in vitro DNA packaging systems have been developed for many of the phages we will describe [T3 (41); T7 (89); T4 (79); λ (50); ϕ29 (40)]. This ability to manipulate DNA packaging has been the hardy complement to the genetic, biochemical, and microscopy approaches that preceded, and now parallel, the development of these experimental systems. More recently, structural data have come to the forefront of efforts to understand DNA packaging in the form of cryo-electron microscopic reconstruction of phage structures and X-ray crystallographic and NMR analyses of components of the DNA packaging machine. These advances bring additional relevance to the study of DNA packaging in bacteriophages and offer the opportunity to elucidate the mechanism of DNA packaging.

Components of DNA Packaging Systems

In order to provide an informative account of the phage DNA packaging process, we will first briefly review the components involved in packaging in some well-characterized phage systems. All these phages have a double-stranded DNA to be packaged; a prohead receptacle for the packaged DNA; and packaging ATPases, enzymes that procure the DNA substrate and mediate the conversion of chemical energy to the mechanical energy required to translocate the DNA into the prohead. The convergence of the maturation pathways and the interaction of these components comprises the DNA packaging event.

DNA

The phage DNA chromosome must retain the information to do three things: ensure its own replication to produce chromosomes to be encapsulated into progeny virions; commandeer the host cell metabolism and redirect it toward the production of progeny virions; and encode the structural proteins and enzymes required to assemble new virions. To achieve the first goal, a number of strategies yield forms of replicated DNA that are presented as immature chromosomes to be packaged. Virion DNA of the dsDNA phages is linear and is packaged processively, generally from left to right with respect to the conventional genetic map

Table 6-1 Viral DNA and DNA Replication Strategies of Various Double-Stranded DNA Phages

Phage	Incoming virion DNA	Replication strategy	End state of replicated DNA
λ	Linear with 5′ 12-base complementary ends (cohesive ends)	Closed circle switching to rolling circle	Linear concatamer
P22	Linear with 104% terminal redundancy	Recombination with extension via direct repeats	Linear concatamer
ϕ29	Linear with covalently attached gp3 at 5′ ends	gp3-primed extension, strand displacement	Unit length with gp3 covalently attached at 5′ ends
T3 (T7)	Linear with 230 (160) base pair direct repeats	Recombination with extension via direct repeats	Linear concatamer
T4	Linear with 102% terminal redundancy	Invasive strand initiated via terminal redundancy	Branched concatamer
SPP1	Linear with 104% terminal redundancy	Unknown	Linear concatamer

(an exception is the T3/T7 systems, which package right to left). There is a teleonomic relationship between the DNA replication strategy of a given phage and the form of the linear DNA encapsulated in the virion. The key to the relationship lies in the replication of a linear DNA molecule upon infection without loss of genetic information needed to prime DNA synthesis at the 5′ end. DNA replication strategies and the resulting structure of the DNA packaging substrates are summarized in table 6-1.

An accessible form of DNA is the defined unit length chromosome produced by ϕ29. Attached to each 5′ end of the 19 kb ϕ29 dsDNA is a covalently linked terminal protein, gene product 3 (gp3) (68). This DNA-terminal protein complex, which is analogous to DNA-terminal protein complex in adenovirus (80), is capable of priming DNA replication from each end, thus providing a straightforward means of overcoming the loss of information in lagging strand synthesis (69). The result is a mature, unit length chromosome that can act as a ready substrate for DNA packaging (6). Similar to ϕ29, unit length DNA is produced during replication of the phage P2 genome. Unlike ϕ29, however, P2 DNA is replicated via a closed circle mechanism similar to plasmid replication (4). In P2 the covalently closed, circular DNA (77) is processed to a linear form for packaging (11, 12). Since this linear molecule must have the capacity to recircularize upon entry into the host cell, the packaging apparatus generates 19-base 5′ overhangs that mediate circularization. DNA replication is not so simple in other phages, however, and the substrate chromosome for DNA packaging rarely appears in such an accessible form.

DNA replication during infection by many well-studied dsDNA phages produces a substrate DNA for packaging that is a composite of individual genome lengths organized into head–tail concatemers. λ circularizes its infecting DNA molecule via the 12 bp sticky ends. Unlike in P2, initial closed circle replication employing a single origin on the DNA is displaced by a rolling circle mechanism that produces DNA concatemers several genomes in length. Thus, to recapitulate the linear chromosome, DNA packaging resolves single copies of the chromosome with the 5′ overhangs from the double-stranded concatemer (see below).

The linear virion DNA of many other dsDNA phage types is, unlike ϕ29, P2, and λ, longer than the length of the genome. In phage Mu, which integrates its DNA into the host cell genome, the additional DNA is of host origin, the result of excision of a length of DNA greater than the length of the integrated phage genome (43). In phages T3 and T7, P22, SPP1, and T4, the linear virion DNA is terminally redundant, with a portion of the DNA sequence at one end of the genome being repeated at the other end of the DNA. This terminal redundancy permits replication without the loss of genetic information in that, although linear replication causes loss of information at the 5′ ends, the redundancy allows the entire sequence to be recovered during subsequent replication (55). These phages employ a variety of mechanisms to generate long concatemers that depend upon this terminal redundancy of the chromosome, which in turn yield terminally redundant genomes during DNA packaging. In some cases (T7, T3), the sequence that makes up the terminal repeat is the same for all virions in the population. In others (T4, P22), the packaging process yields a population of packaged genomes that are circularly permuted with respect to each other and therefore have different terminally redundant sequences in different particles.

While the exact mechanism of replication to form linear concatemers for phages λ, P22, T3, and T7 varies (55), the end result is a packaging substrate consisting of a long molecule comprised of multiple copies of the genome from which a virion's complement of DNA is procured during packaging. An additional complication is faced by phage T4, whose invasive strand replication initiation yields not only long concatemers, but ones containing numerous Holliday junctions that leave them highly branched (23). These convolutions must be resolved during packaging to yield the appropriate linear DNA to be translocated into the phage head. Phage Mu, whose DNA is integrated into the host genome, must excise copies of its genome from the host chromosome prior to, or concomitant with, DNA translocation (43). In each of these systems a complex series of enzymes and processes effect maturation of the DNA substrate and mediate its encapsidation.

Packaging Enzymes

The task of retrieving the phage DNA and processing it to a packagable form rests with a collection of proteins forming a complex often referred to as the terminase holoenzyme (16). This term belies the primary function of these enzyme complexes in many phage systems, where they perform the task of retrieving the unit length DNA packaging substrate from the long concatemers formed by the myriad DNA replication strategies. This definition underrepresents the true capacities of this group of proteins since it describes only one of multiple functions during packaging. In addition to cleaving the substrate DNA to terminate packaging and generate a new end, terminase complexes target the DNA to the waiting prohead and mediate ATP hydrolysis to power DNA translocation, possibly acting as the primary transducer of force during translocation (see below). The designation terminase does not apply to phages which package a preformed unit length genome, such as ϕ29, where the enzyme is more appropriately termed the packaging ATPase.

All known terminase holoenzyme (packaging ATPase) complexes function as a complex of two proteins. The classical terminase combination consists of a large and a small protein, each with specific activities. The small subunit recognizes and binds to specific sequences in the substrate DNA in most phage systems and positions the large terminase subunit to cleave the DNA. Endonuclease activity invariably lies in the large subunit, as well as the ATPase activity responsible for mediating DNA translocation (42, 70). For most systems the DNA-bound large subunit interacts with the prohead, and hydrolyzes ATP to power DNA translocation.

Proheads

DNA packaging culminates in the insertion of the mature DNA substrate procured by the terminase holoenzyme into a receptive prohead (figure 6-1). The icosahedral head shells of the dsDNA phages share the same basic architecture and maturation pathway (45). The major shell protein polymerizes to form the prohead shell by associating with the head–tail connector and scaffold-core components. The dodecameric connector (or portal protein) is embedded at

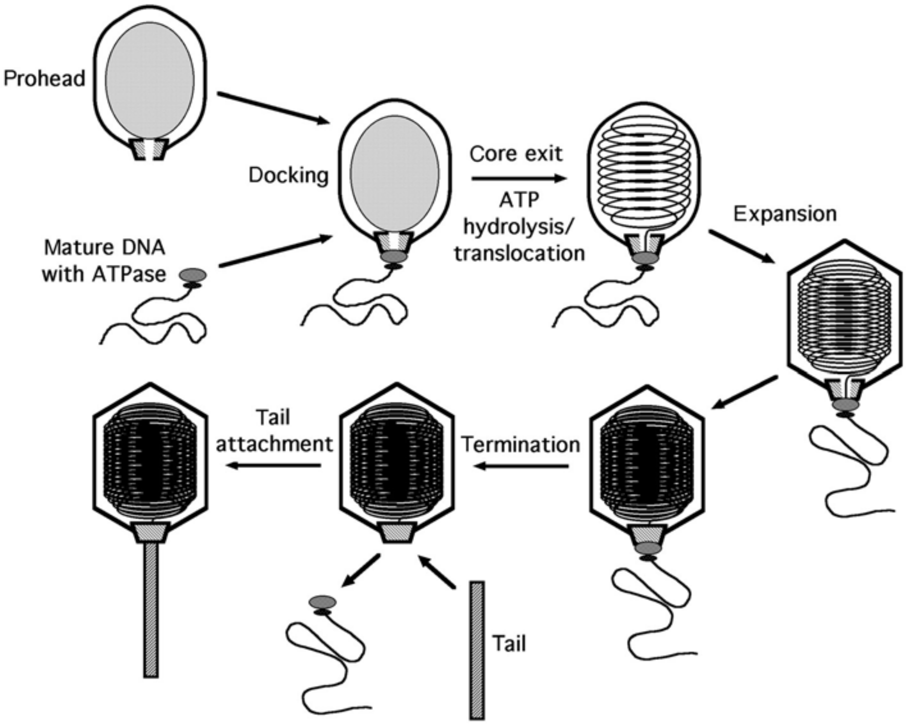


Figure 6-1 Schematic of generalized dsDNA phage assembly. A prohead interacts with the packaging ATPase holoenzyme–DNA complex via its head–tail connector. ATP hydrolysis powers translocation of the mature DNA, and at some point the scaffold core is ejected, either whole or following proteolysis. After an amount of DNA enters the head, the shell capsomeres rearrange, making the head more angular and, in most phages, increasing the head volume. DNA translocation continues until a full complement of DNA enters the head, determined either by the unit length of the DNA, sequence recognition of the DNA length, or a headful mechanism. The ATPase–DNA complex detaches from the connector and is replaced by neck, tail, and/or tail fiber components, yielding a mature, infectious virion.

one of the 12 vertices termed the portal vertex (93). Nomenclature diverges in different phage systems, with the core-containing structure being termed prohead (λ, ϕ29), prehead (T4), or procapsid (P22). For the sake of consistency, we refer to this precursor capsid shell as the prohead.

There is a common maturation pathway of the prohead for most dsDNA phages that varies only in detail (figure 6-1). Putative prohead structures can be isolated from mutants lacking DNA packaging components, and they may or may not contain scaffolding protein. Prohead-like particles may be defective for packaging because they are unstable or immature (27), reflecting a need for synchrony of prohead maturation and packaging. For example, expansion of the capsid and concomitant thinning of its wall that is programmed to occur in DNA packaging may have initiated or occurred prematurely. The role of these structural maturation events has been probed in detail with respect to their mechanistic and temporal relationship with DNA packaging initiation and DNA translocation (see below). In general, the viable receptacle for DNA packaging is the unexpanded, core-containing prohead, with any proteolytic maturation and shell expansion occurring after DNA packaging initiation.

Occupying a unique vertex of the prohead is a multifunctional structure called the head–tail connector or portal that is essential in prohead assembly and DNA packaging (96). The distinction between these two terms lies in the role this structure plays at different times in assembly. The term portal refers to its role in facilitating the passage of DNA into and out of the prohead, whereas the term connector refers to its role as the junction between prohead and tail. We favor the term connector, simply due to its preference in the systems with which we are most familiar (ϕ29, T4). Prior to DNA packaging, the connector plays a role in the initiation of head shell formation by interacting with both the scaffold and head shell proteins. In phage T4, the gp20 connector also interacts with gp40 on the inner surface of the cell membrane to initiate head formation (101). During packaging, the connector binds the mature DNA packaging ATPase complex, is the portal for entry of the DNA (possibly playing an active role in translocation), and is involved in the signaling for packaging termination. Following the completion of packaging the connector is the target for tail assembly, and in the mature virion it has a role in release of DNA during infection. That the connector is capable of engaging in each of these processes in a precise order speaks to its remarkable capacity not only to do many things, but to do them at the right time.

DNA Packaging Processes

In the infected cell viral DNA is recognized by the packaging proteins in a background of host polynucleotides. In spite of differences in the mechanics of DNA replication in different phages, as well as the persistence or absence of an intact host genome, there may be a single mechanism for phage DNA maturation for packaging that is grounded in DNA end formation. DNA maturation for packaging is defined as targeting of the resolved phage chromosome to the waiting prohead. The dsDNA phages select genomic DNA efficiently from the myriad pool of nucleic acids within the cell as evidenced by the high efficiency of infection by progeny, nearing 100% for most dsDNA phages.

Does the DNA packaging enzyme complex pre-assemble and then target the mature prohead, or does it assemble on the prohead? Is this prohead targeting event correlated with prohead maturation events or with DNA replication or transcription? These points may be crucial in that individual events can be temporally or, more importantly, mechanistically coupled to one another.

Once the prohead and DNA are linked and the DNA is positioned for packaging, how is DNA translocated? The structure and mechanism of the motor and the nature of the chemomechanical energy conversion are the areas of greatest current interest and experimental focus. After the complement of DNA has entered the prohead, packaging is terminated. The unit length of the DNA packaged can be measured by targeting a DNA sequence to signal that the head contains one genome, or the amount of DNA in the head may feedback on the packaging machine to trigger termination. By compiling what is known for each phage, our intent is to describe a general DNA packaging mechanism for all dsDNA phages. However, a universal mechanism for DNA translocation might not exist, and caution is needed in comparing individual facts from disparate systems.

DNA Maturation

Maturation of the phage DNA from the cytoplasm of the infected cell is the first event of packaging (figure 6-2). As mentioned, phages such as ϕ29 and P2 replicate unit length chromosomes. Phages λ, P22, SPP1, T3 and T7, and T4 produce long concatemers of DNA comprised of a number of copies of the genome linked head to tail; unit length chromosomes are cut from these concatemers and packaged.

DNA maturation events in phage λ are quite well understood (17). End formation occurs at the structurally complex *cos* site, which spans 200 bp of DNA at the ends of the genome. The terminase holoenzyme complex of the two λ packaging proteins, gpNu1 and gpA, binds *cos* through the interaction of Nu1 with three sequence domains, R1 through R3, of *cosB* ("binding") on the right of the *cos* site. The larger subunit, gpA, then catalyzes a single-stranded nicking reaction in the central *cosN* region ("nicking"), in the center of the *cos* site, producing 12 bp 5′ overhangs. Subunit gpA is thought to bind as a dimer, thus permitting cutting of both strands on one side of the DNA helix to generate the 12 bp

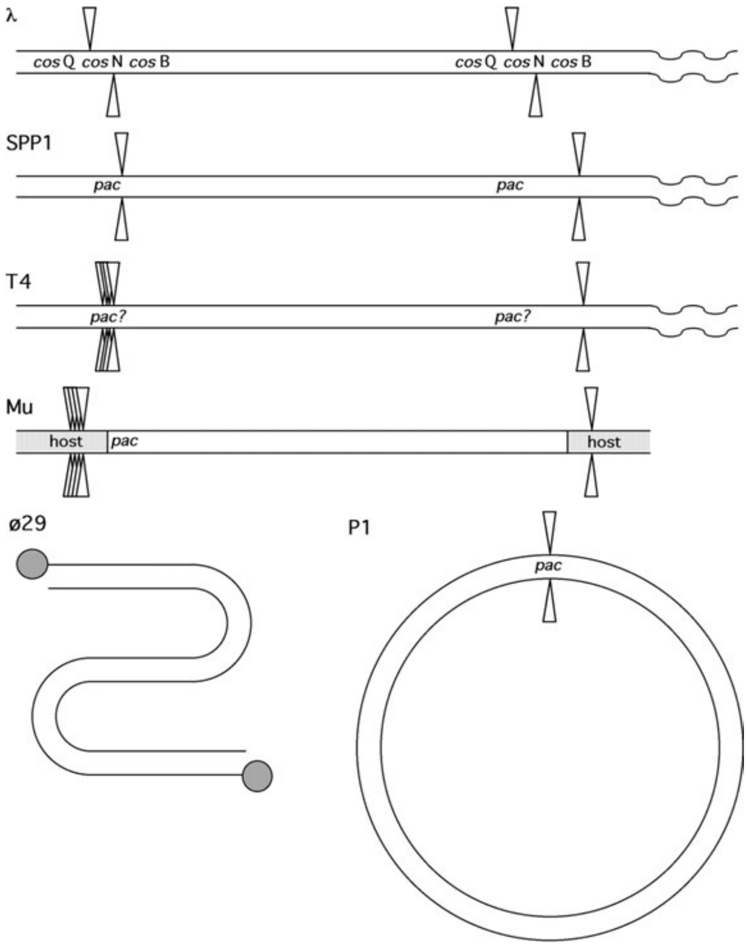


Figure 6-2 Schematic of selected strategies for DNA maturation and packaging. Maturation ranges from formation of defined unit length chromosomes in λ, ϕ29 and P1, to terminally redundant SPP1 and T4. Mu retrieves DNA from the host genome upon excision. Structure of the *pac* site and accuracy/location of cleavage (arrows) with respect to *pac* varies among phages.

overhang. Although gpA alone can bind and cut λ DNA in vitro, apparently Nu1 is crucial for efficient targeting in vivo. Terminase binding and cleavage initiation in λ also involves the action of IHF (integration host factor), which binds the region in the *cos* site between R1 and R2. IHF bends the DNA in such a way that a dimer of Nu1 can bind the juxtaposed R1 and R2 sites (24). Once DNA is cleaved by terminase, strand separation occurs via an ATP-dependent process (42), possibly to rearrange and activate the terminase subunits bound to the DNA for an additional, as yet unidentified maturation process. Terminase preference for the right side of the *cos* site is driven by gpNu1 binding to *cosB* to form the stable intermediate, complex I. The union of complex I with the λ prohead yields the ternary complex II, which then proceeds to translocate the DNA through the connector and into the capsid.

Termination of λ DNA packaging is achieved by sequence recognition of the downstream *cos* site by the engaged packaging apparatus during DNA translocation. The *cosQ* region on the left of the downstream *cos* complex is recognized, and a *cosB* independent cleavage occurs at *cosN* to generate a complementary 5′ overhang on the end of the packaged chromosome (100). *cos* cleavage is sequence-specific but also involves detection of the amount of DNA that has been packaged, since constructs with less than 78% of the normal complement of DNA between *cos* sites fail to cut normally (31). This feedback mechanism plays a role in other phages (see below). The termination of each packaging cycle from a concatemer regenerates the complex I, which can initiate a new translocation event. On average, each termination-initiation cleavage event is capable of priming between two and three translocation events without the need to generate complex I from the DNA concatemer de novo (32).

The other phages requiring a cleavage of their DNA to terminate packaging and to generate a free end for the next cycle display variations of the λ archetype. The terminase complex recognizes and binds a defined *pac* site on the DNA and catalyzes cleavage to generate an end for packaging compatible with requirements for DNA replication on infection. The circular unit length chromosome of phage P2 is cleaved by terminase at a *pac* site to generate the linear DNA to be packaged (11, 12). Like λ, P2 terminase endonuclease generates 5′ single-stranded ends, 19 bases long. P2 DNA cleavage is coupled to prohead docking in that viable proheads must be present in vitro in order for the terminase complex to target and cut the DNA (12, 77). Phages such as T3 and T7, P22, SPP1, and Mu employ a strategy similar to that of λ in that a holoenzyme complex of the two terminase proteins targets a *pac* site. Differences are found in the precision, i.e., the location, of cuts made in the precursor DNA relative to the *pac* site. As in λ, DNA processing accommodates the requirements for DNA replication upon infection. Phages T3 and T7 are proposed to make staggered, defined single-stranded cuts at the *pac* site, but these nicks are 230 and 160 bp apart, respectively. It was originally proposed that DNA synthesis separates the strands between these nicks and regenerates double-stranded DNA with blunt ends from the single-stranded ends, ensuring the terminal redundancy needed for genetic competence of the progeny virion (97). More recently, a double-strand break mechanism has been proposed (34). The terminal redundancy is preserved at the right end via a nicking and replication mechanism in which a displaced template is produced, followed by a double-stranded cut that retrieves the right end of the packaging genome from the concatemer. Packaging terminates with a double-stranded cut at the left end. An analogue of involvement of IHF in λ packaging appears to be the bending of SPP1 DNA mediated by the gp1 small terminase subunit (19).

Phages P22 and Mu maintain a *pac* site that is targeted by their respective terminases, but cleave the DNA nonspecifically (Mu), or semispecifically (P22) in a region of the adjoining DNA. In the case of Mu, whose unit length genome is integrated into the host cell chromosome, the initiation cut is upstream of the phage *pac* site (35). Therefore Mu retrieves a small portion of the host chromosome DNA, on the order of 56–144 bp, on the left end of the phage DNA to be packaged. P22 makes an initiation cleavage within a target area of approximately 120 bp of its *pac* site, generating a blunt end DNA capable of packaging initiation (3, 14).

In addition to the relative lack of fidelity in initiation cleavage, Mu and P22 do not terminate packaging at a predetermined sequence as in λ or T3 and T7. Rather, these two phages engage in a headful packaging mechanism in which the sequence-independent cleavage of the DNA is determined by the amount of DNA packaged. Packaging of more than one genome length of DNA ensures replication competence upon infection. In Mu, host sequences to the right of the phage genome are packaged (20). In P22, 104% of the genome is packaged, providing the terminal redundancy needed for DNA replication (13). How termination cleavage is triggered is unknown, but the physical force of the compacted DNA against some component of the packaging machine, either the connector or the ATPases, may signal that the head is full. Work on P22 and SPP1, which use headful packaging, has demonstrated that mutations in the connector affect the length of DNA packaged (15, 94). These mutants suggest that headful packaging control lies in the connector but leave open the possibility that the terminase complex engaged in packaging could be altered indirectly by these mutations. In SPP1 and P22, as in λ, the termination of the initial DNA packaging event regenerates the initiation complex that can target the next available prohead. Unlike λ, however, the left ends of the DNA generated from subsequent rounds of packaging are staggered downstream in increments of 4% (in P22) or 5.6% (in SPP1) of the genome length as a result of the headful mechanism described above.

Phage T4 is unique in comparison with other well-studied dsDNA phages in that it does not have a defined, sequence-specific *pac* site. The holoenzyme complex of the large T4 terminase protein, gp17, and the small terminase protein, gp16, binds the hydroxymethylated T4 DNA via gp16 targeting. While no particular sequence is recognized, initiation cleavage was recently shown to be coupled to recognition of single-stranded regions generated by replication initiation and transcription (33). gp17 has a domain for binding to single-stranded DNA, and binding is augmented by the smaller gp16 terminase subunit. Therefore, while not sequence-specific per se, this requirement for single-stranded DNA suggests that initiation cleavage in T4 is not entirely a random event since it is coupled to sequence-specific processes. The large terminase subunit, gp17, interacts directly with the connector, gp20 (62), and probably the ATPase activity of gp17 powers both DNA cutting and translocation. Recently it has been shown that T4 gp17 also interacts with the phage late sigma factor gp55 (67), implying that the gp17 terminase subunit targeting of the DNA is in part directed by a cofactor similar to those seen in λ and SPP1 (see above).

As mentioned, T4 DNA packaging must also resolve a large number of branches in the substrate DNA in order to produce an intact linear DNA genome. Endonuclease VII, gp49, is responsible for much of this work, although gp17 alone might be capable of resolving most branches since some filled heads are produced in the absence of gp49 (65). The gp49 works as a resolvase capable of trimming the branched DNA at the Holliday junctions left over from invasive strand-initiated replication (71). Originally it was thought that gp49 acted downstream of the packaging complex on the prohead, but more recently it has been shown that this enzyme is in contact with the gp20 connector during packaging (36). The gp49 binds a discrete domain of the connector and is not in contact with the gp17.

T4 DNA packaging is terminated by a headful packaging mechanism, with cutting being mediated by gp17. Isometric, petite heads package DNA to the same density as larger prolate heads, yielding a virion chromosome 40% smaller than normal (30). In addition, canavanine- or head shell mutation-induced lollipop monster phages are capable of packaging large DNA molecules in the megabase range (22).

The unit length ϕ29 DNA with its associated gp3 terminal protein is competent for DNA packaging after replication, without the need for cutting seen in other phages. However, a complex series of steps generates a DNA substrate capable of interacting with the prohead. Sequence-independent interaction of the terminal protein, gp3, with the downstream DNA helix has been described (37). This produces lariat loops that can form without the participation of other phage proteins. Only one gp3 is required, since lariats can form from various lengths of either right- or left-end fragments of the DNA generated by restriction endonuclease digestion. Binding of the gp16 packaging ATPase to gp3 at the lariat loop junction permits the introduction of supercoils into the lariat. Whether this supercoiling event is ATP-dependent is unclear, and the mechanism by which the DNA is twisted has not been resolved.

This higher order conformation of the DNA–gp3–gp16 seen in ϕ29 likely provides efficient packaging initiation. It has been demonstrated that free ϕ29 connectors wrap about 1.6 turns of DNA around their outside surface (95), and gel electrophoresis and electron microscopy show that the connector embedded in the intact prohead also binds and wraps supercoiled DNA (D. Anderson, unpublished data). This event provides a mechanism by which the end of the DNA is targeted to the connector portal to initiate packaging. Considering that the end of the mature chromosome is a relatively small target in the busy environment of the cell, the long axis of the DNA is a large target for any complex capable of recognizing and interacting with the DNA. We suggest that ϕ29 takes advantage of this by targeting the long axis of the DNA and then moving in one dimension to the end to initiate translocation. Whether nicked DNA, being torsionally unconstrained, can initiate packaging has not been clearly resolved. It is also unknown whether such DNA tertiary structure is employed by other dsDNA phages as a means of targeting the DNA packaging enzyme complex to the prohead.

A second possible function for the wrapping of supercoiled DNA around the connector is to effect connector conformational change that converts it from a static organizer of shell assembly to a dynamic packaging-motor organelle. This idea is based on the finding that wrapping of supercoiled plasmid DNA around the ϕ29 prohead-embedded connector allows the shell, previously tightly fixed to the connector, to be easily stripped away. The contour length of the DNA–connector complex that remains is reduced by about 120 bp, suggesting that DNA wraps around the connector and restrains a negative supercoil, as demonstrated previously for the free connector-supercoiled DNA complex (C. Peterson and D. Anderson, unpublished data; 95).

Prohead Maturation

A common theme in dsDNA phage head assembly is the maturation of the prohead from a fragile, scaffold/core-containing precursor to a stable mature form. The head shell likely polymerizes around a scaffold–connector complex in a relatively unstable form, which later transforms to a rigid and structurally robust capsid (27). This structural conversion can involve proteolytic cleavage of the scaffold and/or shell proteins and an increase in the prohead volume by as much as 100% (HK97) (21), a process called expansion. Scaffold exit may precede or occur concurrently with expansion.

While both scaffold exit and capsid expansion (with the exception of ϕ29, which does not expand) must occur to allow the full complement of DNA to enter the prohead, the question persists of whether they contribute directly to DNA translocation (see below). Considering scaffold processing, early models of DNA translocation focused on the compacted phage chromosome as an analogue of the condensed DNA produced by polyvalent cation-mediated collapse of DNA into a toroid (59). In some phages, such as T4, it was suggested that core cleavage might produce small, charged peptides capable of condensing the phage chromosome within the head, thus drawing the long linear DNA through the portal vertex (59). Studies on phages T7 and P22 suggest that the core exit may be coupled to DNA packaging, since only core-containing particles are packaged in vitro (57, 81). In vivo observation of T4 offered additional support in that only core-containing particles could be chased into phage during both wild-type and mutant infection (5, 60, 64). On the contrary, λ proheads can package DNA after core exit (46), and ϕ29 proheads containing only about five copies of the normal complement of about 150 copies of scaffolding protein are packaged efficiently in vitro (D. Anderson, unpublished data). These conflicting observations have not been reconciled, and there is no current mechanistic model relating scaffold exit and DNA packaging.

Is the structural transformation of the lattice in prohead expansion mechanistically linked to DNA translocation? Only unexpanded proheads of λ, T7, and P22 can package DNA in vitro, and expansion occurs during packaging (27, 86). This led to the suggestion that prohead expansion might drive DNA translocation. Decrease in ion permeability of the head shell during expansion in T7 prompted the hypothesis that the DNA might be sucked into the sealed prohead by the hydrostatic pressure created during expansion (83).

However, though increase in head shell volume is dramatic, for example, on the order of 50% in T4, this increase is insufficient to account for the amount of DNA translocated into proheads exhibiting capsid expansion (27). DNA translocation is ATP-dependent for all phages, and no link between prohead expansion and ATP hydrolysis has been described other than the synchrony of packaging and expansion. In addition, ϕ29, which exhibits capsomere structural change in packaging, does not show a detectable increase in shell volume (93). In phage T3, which initiates packaging into unexpanded proheads, expansion occurs discretely after a small portion of the DNA complement is translocated (86). Similarly, the capsomere change in ϕ29 probably occurs after only a few hundred base pairs of the DNA enter the prohead (7). In T3, packaging can be stopped in vitro after capsid expansion by the addition of ATP analogs such as γ-*S*-ATP, and then restarted and completed into the expanded prohead upon restoration of ATP (86). The hydrostatic pump model has been revised such that a regenerated hydrostatic pressure is maintained across the prohead shell which pulls the DNA into the capsid (85). This model still suffers, however, from the observations that effective translocation can occur after initial prohead expansion (79, 86) and that expansion is most likely irreversible (91).

The singular test case that unlinks expansion and DNA packaging is the report of in vitro packaging of a phage T4 particle after both core cleavage and expansion (79). However, T4 DNA packaging in vivo cannot occur after scaffold cleavage and shell expansion (63). Temperature shift experiments with temperature-sensitive and cold-sensitive mutants in the T4 terminases show that the expanded prohead is not rescued in vivo. The only proheads that have not initiated packaging that can be rescued in similar experiments are scaffold-containing, unexpanded proheads (53), implying that after scaffold cleavage, the prohead proceeds down a defective pathway in vivo. To credit the in vitro experiments requires the assumption that expanded proheads can be rescued only with transfer into the in vitro world. Additionally, expanded T4 proheads assemble tails in vivo without packaging DNA during infection with terminase mutants (52). This suggests strongly that expansion prior to DNA packaging is aberrant. It has been shown recently that the normal in vivo substrate for DNA packaging initiation in T4 is the unexpanded prohead, and that prohead expansion occurs after a significant amount of DNA enters the prohead, as in other phages (51). The newly reported high-efficiency in vitro T4 packaging protocol (67) can be used to retest the activity of the expanded prohead.

If DNA packaging is not mechanistically coupled to head maturation events, what is the impetus for scaffold exit and prohead expansion in DNA translocation? The answer may lie in subtle events of packaging initiation and early DNA translocation that involve connector conformational change and, consequently, irreversible conformational change in the shell. Ordered virion assembly is ensured by binding the DNA packaging apparatus to an unfilled prohead, but not a filled particle. Thus, DNA packaging is coupled temporally to core cleavage and prohead expansion so that packaging precedes tail attachment.

How DNA packaging triggers prohead expansion is unknown. Expansion of the T4 polyhead lattice in vitro under conditions of low salt is unidirectional and exothermic (90, 91). It is hypothesized that enough DNA to form a single layer within the prohead, contacting the inside of the head shell, might trigger expansion. However, apparently much less packaged DNA, on the order of a few hundred base pairs, is thought to trigger capsomere conformational change in ϕ29, albeit without expansion. Expansion might also play a role in ensuring proper organization of the packaged DNA by opening new binding sites on the inner surface of the capsomeres. Possibly changes in the connector may propagate a wave of expansion up the head. Connector conformational changes likely mediate packaging initiation, establish the sensing mechanism for headful packaging in some phages, and cue ordered tail attachment. A certain consequence of shell expansion is to provide stability to both the nascent and the mature virion.

The Mechanism of DNA Translocation

Union of the mature substrate DNA and the ATPase (terminase) holoenzyme complex with the viable prohead results in the assembly of a molecular machine that is capable of translocating DNA into the prohead. After decades of effort, the exact mechanism of DNA translocation is unknown. Key in the search for the mechanism of DNA translocation is that ATP hydrolysis is the driving force behind DNA translocation in all in vitro systems. Moreover, all identified DNA packaging holoenzyme complexes have the capacity to hydrolyze ATP. Therefore, the task is to define where, when and how ATP hydrolysis is used to move DNA around or through the connector and into the head.

First, and most poorly understood, is the mechanism by which the free end of the double-stranded DNA substrate, once engaged with the head–tail connector, is introduced into the portal pore. While DNA deposition may involve a continuous transfer from the engaged ATPase complex through the connector pore, this delicate initiation must depend on the highly evolved fidelity of the terminase holoenzyme and connector.

The mechanism of DNA translocation relates broadly to how molecular machines in general harness biochemical events to achieve movement of molecules. Included in the list of well-studied motors are the myosin ratchet, the F1-ATPase rotary motor, and the RNA and DNA polymerases. The same questions that persist in these motors apply to the DNA translocation motor: Is there a bias to trap favorable Brownian motions by a sequence of small free energy drops (power stroke), or is a run of favorable thermal fluctuations rectified by a large free energy drop (Brownian ratchet) (75)?

At the center of past efforts in describing the mechanism of DNA translocation lie the ATPases themselves. The simplest, and most appealing, mechanism is one in which the terminase holoenzyme plays a direct role in translocation. It has been suggested that the oligomeric ATPase

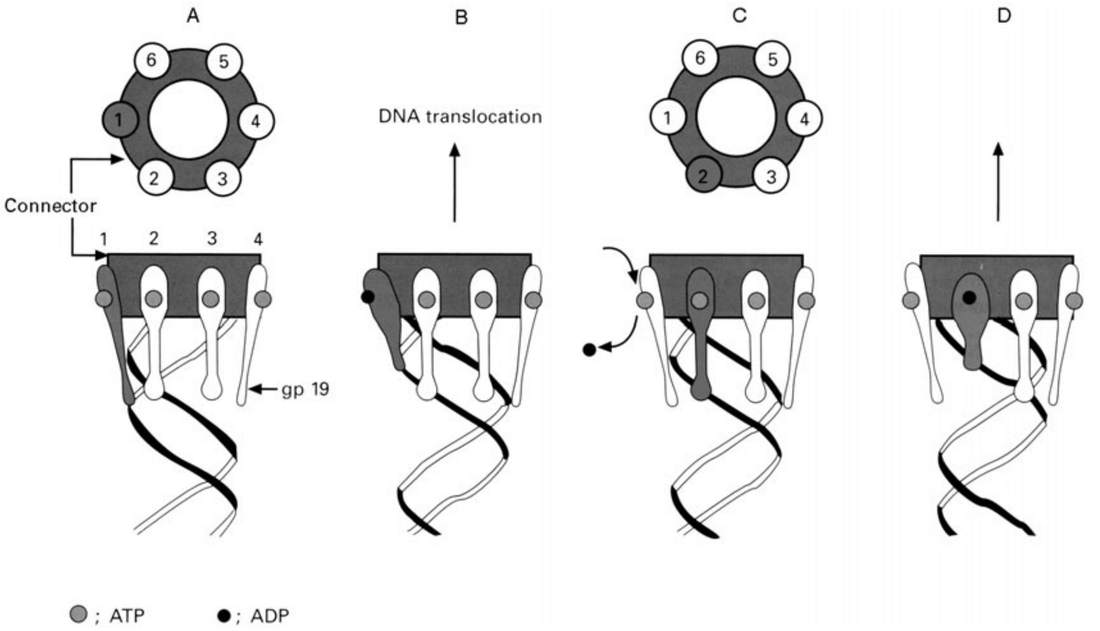


Figure 6-3 An ATPase ratchet model of DNA translocation in T3. A: gp19 ATPase subunit positioned on the connector interacts with the phosphate backbone of the DNA. B: Upon hydrolysis of ATP, the gp19 monomer changes conformation, driving the DNA through the connector and into the prohead. C: This translocation event brings the backbone into alignment with the next consecutive ATPase monomer. D: the next translocation event is initiated. Reproduced from Fujisawa and Morita (34) with permission.

holoenzyme associated with the connector during DNA packaging might be able to oscillate up and down with respect to the axis of the DNA entering the connector (34). In a mechanism similar to the myosin head ratchet, the ATPase subunits would "walk," either individually or in concert, unidirectionally along the DNA helix, and the DNA would be translocated in the process (figure 6-3). The precise nature of the structural transitions in the ATPase complex that are required to fulfill this mechanism have yet to be described.

Similarly, the ATPase holoenzyme complex could directly translocate the DNA into the prohead via a mechanism similar to polymerase tracking along the DNA. If the ATPase complex is fixed at the connector portal and creeps along the DNA like a polymerase, the result would be translocation. This polymerase-creeping model and the similar ATPase ratchet model draw support from the inhibition of packaging by DNA intercalating compounds and the detected ability to package DNA with gaps or nicks. This implies that the exterior phosphate backbone of the DNA is the point of interaction between the translocating motor and the substrate (34). Whether the ATPase holoenzyme plays a direct role in energy transduction or not, the hydrolysis of ATP catalyzed by the terminase complex plays a defining role, as evidenced by the effect of certain mutations in the ATPase region of the λ gpA protein on the rate and efficiency of translocation in vivo (26).

Many current models of DNA packaging hypothesize a role for the symmetry mismatch between the dodecameric connector and the 5-fold symmetric vertex of the icosahedral shell in which the connector is embedded (figure 6-4a). This symmetry mismatch potentiates rotation of the connector within the prohead shell (44) by abrogating the rigid interaction of components of like symmetry. The symmetry mismatch of the connector and capsid is confirmed by cryo-electron microscopy three-dimensional reconstruction of ϕ29, which also shows that the connector appears to fit loosely in the shell (93). Models have been put forward in which connector rotation either actively drives packaging or passively facilitates packaging.

The original connector rotation model of packaging has the helical DNA being driven into the capsid, with either active or passive rotation of the connector, as a bolt passes through a rotating nut (figure 6-4b) (44). It is not clear how an active screw model would satisfy certain requirements of this mechanism, such as the need to axially restrain the DNA to prevent it from being rotated by the connector. Later, the observation that the ϕ29 connector could wrap supercoiled DNA prompted a model in which a rotating connector would move the externally wrapped DNA relative to the head shell (95). This physical displacement could be harnessed in a number of ways to produce DNA translocation, including direct translocation into the prohead similar to a ship's capstan. Alternatively, twisting of the DNA causing the introduction of supercoils into the DNA by the rotating connector or by a packaging ATPase activity (10) could put strain on the DNA so that it enters the prohead to eliminate this superhelical stress. As yet there is no idea

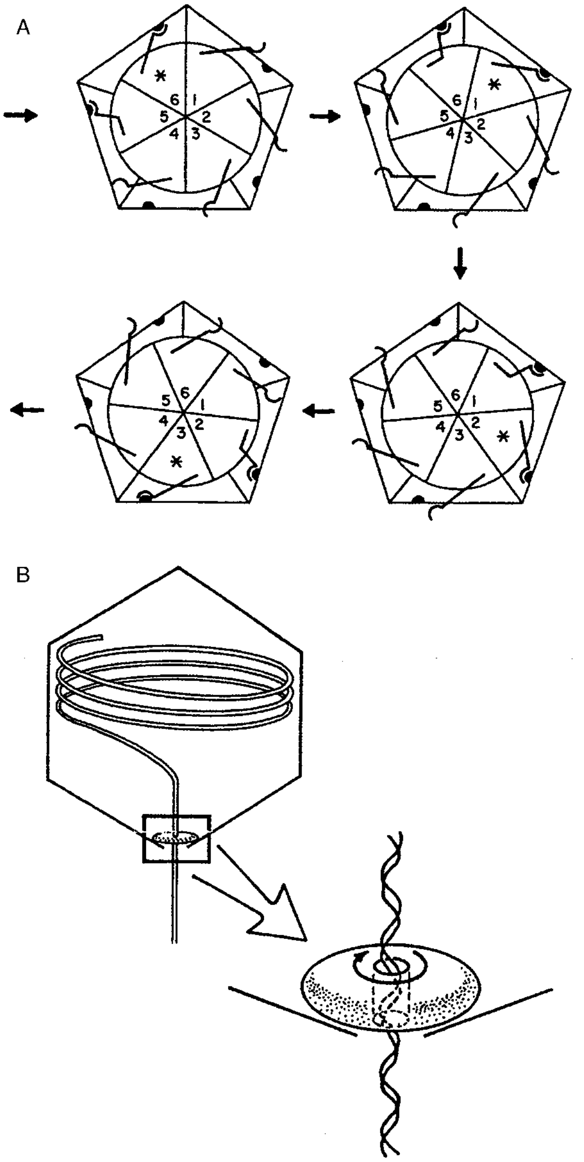


Figure 6-4 Symmetry mismatch between the prohead and connector potentiates connector rotation and DNA translocation. A: Due to the 5–6 mismatch, only a single vertex of each component is aligned at any one time, thus facilitating rotation. If the contact point becomes a lever for translation of the connector with respect to the prohead (*), rotation can be driven by successive events around the subunits. B: If the connector acts like a nut and the DNA helix like a threaded bolt, connector rotation could displace the DNA into the prohead. Reproduced from Hendrix (44) with permission.

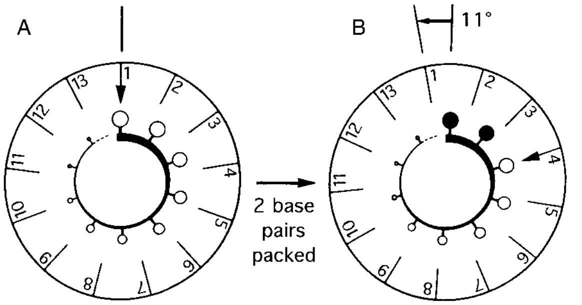


Figure 6-5 Translocation of DNA based on connector rotation and symmetry mismatch between the connector and the DNA in SPP1. A: Contact (arrow) between a monomer of the 13-fold connector and the phosphate backbone of the DNA helix is broken and the connector rotates 11° counterclockwise. B: Contact between the connector and the helical DNA is re-established three connector monomers to the right, two base pairs down the helix, with a consequential translocation of these two base pairs. Reproduced from Dube et al. (25) with permission.

of how movement between the connector and shell is mediated in the active connector rotation models, and connector rotation has not been detected in any phage system. The recent atomic structure of the φ29 connector has generated a model of the packaging mechanism that combines the ratchet and rotation models (see below).

A model that introduces a caveat to the active ATPase models above is one in which the head–tail connector of the phage might move along the DNA backbone in order to achieve DNA translocation. In rationalizing the observed 13-fold symmetry of the free SPP1 connector, Dube et al. (25) proposed a mechanism of translocation in which monomers in the 13-fold portal interact in set sequence with the near 10-fold symmetric DNA helix. As monomers in set sequence bind the backbone of the helix, the DNA must be drawn into the prohead as the perpendicular alignment of the connector and DNA is maintained (figure 6-5). This model was based on a 13-fold model for the SPP1 connector, which has since been shown to be a dodecamer in the prohead like other dsDNA phage connectors (66). However, the 12-fold nature of the connector does not exclude this model.

The φ29 connector structure reveals a most interesting motif. Each of the 12 monomers spans the 7.5 nm high connector from top to bottom (39, 87). However, rather than simply traversing the connector, each monomer has three nearly parallel alpha helices that are canted at an angle approaching 30° to the axis of the connector, giving the overall structure the appearance of a spring. The structure of the connector therefore gives the impression that it is compressible. While no comparative structures have been presented to show connectors of different heights, atomic force microscopy has revealed that the connector can be reversibly compressed by 2.5 nm, about one-third of its height, under loads of 100 piconewtons or more (73). The connector structure and this demonstrated compressibility

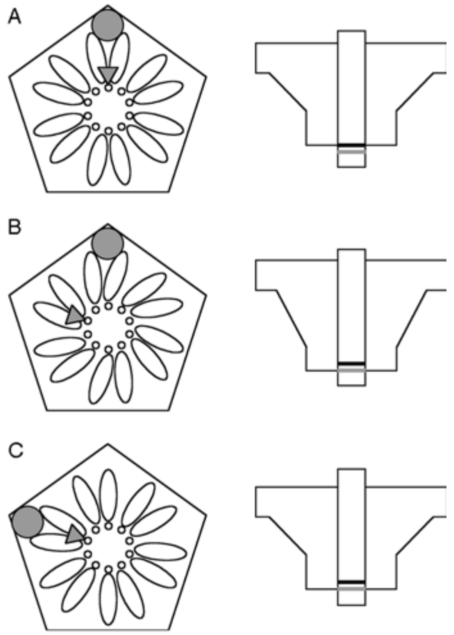


Figure 6-6 Schematic of the compression ratchet model of the packaging mechanism. A: A packaging stroke begins with the alignment of the wide end of the connector with the prohead (gray circle) and the connector channel with the DNA helix (gray triangle); the connector is in a compressed form (right). B: As the connector narrow end releases the DNA (the wide end holds the DNA) and extends, it rotates by 12° counterclockwise with respect to the head such that contact with the DNA shifts to the next pair of connector monomers (left) and two base pairs down the DNA helix (right). C: During the subsequent compression of the connector, the DNA, axially restrained at the narrow end (but released at the wide end), is driven two base pairs into the head (right); concurrently, the wide end of the connector rotates passively 12° counterclockwise with respect to the head (left), re-establishing contact with the head two connector monomers to the left. Reproduced from Grimes et al. (38) with permission.

serve as the basis for the translocation model described below.

It is proposed that the connector oscillates, extending and contracting along the long axis of the DNA inserted in the connector channel (figure 6-6) (87). There are two primary contact regions between the connector and DNA, at the connector wide and narrow ends, respectively. To start a cycle, the DNA is released by the connector narrow end, which rotates by 12° (counterclockwise as viewed toward the head) to maintain contact with the DNA backbone as it extends down the DNA helix by 2 bp. Then the narrow end closes on the DNA as the connector contracts to drive the DNA into the head, and concurrently the connector wide end releases the DNA and rotates 12° to realign the connector with the prohead, pRNA, and gp16 ATPase. These components are reported to be in contact and possess 5-fold symmetry as suggested by cryo-electron microscopy. This cycle repeats. The connector rotation involved is passive and serves to maintain alignment between the 6-fold connector and 5-fold DNA. How ATP hydrolysis mediates these events is unknown, and no direct quantification of the connector dynamics involved has been reported.

Structure of Packaged DNA

Whatever the mechanism of DNA translocation, it must overcome the energetic barrier of compacting the DNA and deliver the DNA into the prohead to confer the proper structure and organization within the head. DNA packaging in dsDNA phages is endothermic. Analogies are often made between DNA packaging and the collapse of DNA into a torus, or DNA toroid, in the presence of polyamines such as spermidine (29) or hexamine cobalt (98). While the final structures share structural similarities, such as the organization of DNA in hexagonal bundles, the processes are very different. DNA condensation by polyamines or hexamine cobalt is spontaneous and exothermic, while DNA compaction in phages requires the input of energy mediated by enzymatic function of the packaging machine. The DNA toroid is stable, while the packaged DNA in phages is metastable. This distinction is crucial, since the function of the phage packaged DNA is to await delivery into a host cell, and it must be ordered in such a way as to permit disassembly of the structure during infection, unlike the DNA toroid which may be irreversibly condensed. Therefore, DNA packaging is defined as a compaction event rather than a condensation.

Several facts seem incontrovertible in describing the structure and organization of the DNA in the phage head. The DNA is in B-form (2), with spacing on the order of 2.5 nm (92). Regardless of the specifics of the overall organization of the DNA, it would appear that it is locally associated in a hexagonal packing array, giving the DNA a quasi-crystalline appearance. What has been debated significantly is the overall organization of the DNA within the head shell. Models proposed over the past four decades describe the gross features of packaged DNA as a wrap solenoid, a liquid crystal, a spiral fold or a folded toroid (figure 6-7). Until recently, insufficient, and often conflicting, experimental data existed to distinguish among these models. However, mounting evidence pushes consensus in the direction of one of the original models of DNA organization in the phage head: the DNA solenoid.

The model for packaged DNA structure that seemingly has the least amount of organization is the liquid crystal model (61). In this model, as in others, the bulk of the DNA is in tightly packed crystalline arrays. These arrays are small, however, and persist within the phage head as discrete domains with local structure, joined to other

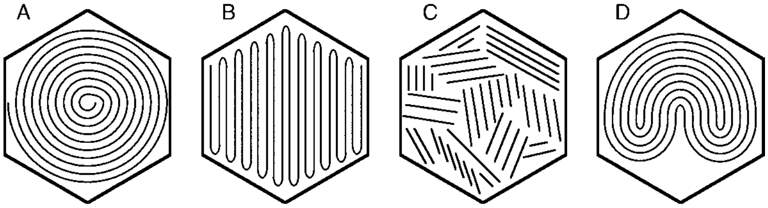


Figure 6-7 Schematic of sectional views of models of DNA packing in dsDNA phage virions. All proposed models of DNA packing include hexagonally packed DNA, and differ according to the extent and type of global organization within the head shell. Models for global organization include: A: the solenoid, B: the spiral fold, C: the liquid crystal, and D: the folded toroid.

randomly arranged packets of ordered DNA by short stretches of disordered DNA. The DNA becomes organized in this fashion as more and more DNA is translocated into the prohead, and the DNA condenses in small regions into hexagonally packed crystals. This is an intuitively appealing model, but much of the experimental observation regarding DNA structure within the phage head appears to be in conflict with this model.

Microscopy studies suggest there is higher symmetry to packaged DNA beyond the level of hexagonal packing promoted in the liquid crystal model. DNA released from disrupted heads often appears as a large coil, suggesting a gross organization of the whole chromosome (28). Cryo-electron micrographs of several phage heads and similar structures reveal patterns in the compacted DNA which resemble fingerprints, suggesting a pattern of loops of DNA inside the head (18, 82). Three models considered below that describe gross organization of the DNA in the head have experimental support. What is crucial, however, is that the model accounts not only for how the DNA can be accommodated in the space of the prohead shell, but also for the DNA translocation event.

One model for packaged DNA is based on a derivative of the condensed toroid (48, 49). The DNA enters the prohead and forms a large donut-shaped structure of hexagonally packed DNA with a hole in the middle, essentially a DNA toroid. As the length of the DNA inside the prohead increases, the toroid collapses into a folded structure, and the packaged DNA density reaches the level found in the mature head. But this attempt to relate the structure of the toroid with packaged DNA fails in that the diameter of the torus that collapses into the folded form inside the head is much greater than the diameter of the prohead shell itself. To arrive in this final form, the DNA must organize into the folded toroid from the outset, or a smaller torus having the size of the head must rearrange into a larger toroid ring of the diameter of the final folded form. These events are unlikely, especially given the requirements for gross rearrangement and the energetically unfavorable sliding of DNA. Thus, while this model may have merit in describing how DNA can fit into the prohead, it does not address the obligate process of how DNA will reach this form during DNA packaging.

A second and earlier model that was derived from experimental observation is the spiral fold model (8, 9). Like others, this model describes hexagonally packed DNA, but unlike the liquid crystal model, the entire chromosome is organized. Briefly, the DNA is arranged as a bundle of straight rods formed by the up and down winding of the DNA along the long axis of the prohead, with the DNA bending back on itself repeatedly, making 180° turns. Several lines of evidence support this model. A series of cross-linking experiments in lambda in which *bis*-psoralen agents are used to interrogate the detail of interaction between packaged DNA and the phage head shell suggest that the DNA contacts the head shell every several hundred base pairs (99). Thus, the kinked DNA at the end of each spiral fold contacts the shell. Ion etching experiments on λ (9), capable of probing the spatial arrangement of DNA within the virion, also support this model.

The spiral fold model, like the folded toroid, presents a reasonable form for the DNA in the full head but is counterintuitive with the nature of DNA translocation. The sharp 180° bends in the DNA described by the spiral fold require the DNA helix to melt at these points. Considering that the DNA is processively driven into the head, it is difficult to imagine what would force the first complement of DNA to align in this way with such severe distortion of the helix. Seemingly the DNA would prefer to trace a path around the sphere of the prohead rather than reverse direction and fold back on itself, since doing so would deny the high persistence length, charge repulsion, and entropic nature of DNA. Recently, new in vivo intracapsid DNA cleavage experiments (72) have led to a slight revision of the spiral fold, bringing it more in line with the solenoid packing model (below).

A third model of packaged DNA, the solenoid, has received ongoing attention for four decades. This model has been reinvigorated by recent support derived from cryo-electron microscopy, which is capable of revealing relatively fine detail without the impediment of artifacts generated by fixation or staining seen in traditional transmission electron microscopy. In most phages, raw images of filled heads or virions reveal a characteristic fingerprint pattern. Cerritelli and others (18) exploited the ability to preferentially orient phage T7 heads in vitreous samples, revealing that all T7 heads have this characteristic pattern. By processing images of these oriented head particles and by regenerating similar images using a theoretical model of a DNA solenoid, some of the best evidence is provided that DNA in phage heads is organized in a layered spool. Assessment of the DNA secondary structure of T7 and other phages by RAMAN spectroscopy is also seen to support this model (76), as does recent theoretical work (see below).

As mentioned above, the key consideration in exploring the structure of packaged DNA is the nature of DNA itself. First, DNA is stiff and has a relatively long persistence length: DNA in solution does not bend back on itself over a distance of less than 50 nm. Second, the highly charged phosphate backbone of DNA makes it self-repulsive: in order to compact DNA to a spacing of 2.5 nm and within the confines of the phage head, this charge repulsion must be overcome and will play a role in determining the organization of packaged DNA. Considering these conditions and the rules of entropic confinement allows for a physical reconstruction of packaged DNA that supports the solenoid model.

Providing more detail, when DNA enters the prohead, its stiffness and self-repulsive nature dictate that it remain relatively straight over the short distance from one side of the prohead to the other. When a length of DNA equivalent to several lengths of the prohead has been translocated, how will the DNA respond? It is likely that the DNA will form loops within the confines of the prohead, following the longest path it can around the inner surface of the prohead shell. As more and more DNA enters the prohead, concentric shells of DNA form, pushed outward from the center of the prohead, driven by the persistence of the DNA. As these layers form, charge repulsion between strands pushes back sequentially from layers at the outside of the shell. Thus the properties of DNA itself are enough to confer some level of higher order structure of the packaged DNA in that a stable equilibrium forms between the persistence of the DNA, which pushes the DNA away from the center of the prohead, and charge repulsion that pushes back. The model of such a solenoid structure formed in this fashion is one of the oldest proposed for packaged phage DNA, but only recently has the model been dealt with theoretically to a suitable degree.

Odijk (74) has calculated the spacing prescribed by such a model and compared it with the observations made by Cerritelli et al. (18). It appears that the spacing observed in T7 heads containing different amounts of DNA agrees well with the theoretically derived spacings based on the principle of equilibrium between DNA stiffness and self-repulsion. Kindt et al. (56) took this theory one step further in an attempt to replicate the way in which DNA is organized in the prohead during translocation. Their dynamic model interrogates how DNA organizes itself during translocation and not simply after the entire DNA complement has entered the prohead. This effort reveals that, at first, the DNA is disordered inside the prohead shell, but soon adopts the conformation of the outer layers of a solenoid. As more and more DNA enters the confines of the head shell, concentric layers form from outside to inside of the solenoid. Although the size of this theoretical prohead deviates from real phage systems by several-fold, the principle supports the idea that the physical nature of DNA alone can organize the DNA within the prohead.

The end state of the packaged DNA also relates directly to the mechanism of DNA translocation in another way. For example, if the final conformation of the DNA inside the capsid shell is a solenoid, then as the DNA enters the prohead through the connector it must rotate axially with respect to the prohead as the incoming DNA winds in concentric coiled rings. A variant of the solenoid proposed by Serwer (84) abrogates this necessity in that it was proposed that rather than spooling continuously in one direction, the DNA reverses direction occasionally in its path around the solenoid. If no such reversals occur in the solenoid, then the translocation mechanism must accommodate the axial rotation required. The rotation of the prohead connector potentiated by the symmetry mismatch between prohead shell and connector leaves open the possibility that, regardless of the details of the mechanism of translocation, such required DNA rotation can be accommodated. Such a rotation event could also occur, or might be necessary, during DNA ejection during infection. In a reversal of packaging translocation, the connector and attached tail components might rotate relative to the head while the DNA moves into the host cell.

The final structure of the DNA brings up an additional point: how much energy is invested in the DNA? This is relevant to the energetics of DNA translocation in that the packaging machine must maintain the capacity to drive the DNA into the prohead. Recent single-molecule optical trap studies with the ϕ29 system have revealed the force–velocity relationship of the packaging motor (88). The DNA packaging machine is remarkably strong at the molecular level, having a maximum stall force on the order of 70 piconewtons. By comparison, this is 5 times stronger than the classical myosin molecular motor (58). Why is the packaging machine so powerful? Force–velocity measurements reveal that the last portion of the ϕ29 chromosome enters the prohead against an internal force of 50 piconewtons. This suggests that the pressure of the packaged DNA within the prohead, and thus the force opposed in translocating the last segment of DNA, is on the order of 6 megapascals. Previous studies suggest that the containment pressure of the packaged DNA might assist in ejection of the DNA into the infected host cell. The ϕ29 work suggests that a significant amount of energy is available for such a process. However, the ϕ29 single-molecule studies do not discern whether all the energy consumed by the packaging machine is deposited in the packaged DNA, or whether some energy is dissipated. Theoretical estimates for the stored energy of the packaged DNA vary, and it is not clear whether the experimental set-up of the optical trap affects the packaging motor. However, the internal capsid force estimated in ϕ29 agrees quite well with the force calculated for expulsion of T7 wild-type versus deletion mutant DNAs in titration calorimetry (78).

Conclusions and Future Considerations

The efforts and information described above are directed toward a single goal: the complete understanding of the processes and events of DNA packaging in dsDNA phages. While much progress has been made, many questions remain. The elucidation of the mechanisms involved in DNA packaging can be tackled like any other molecular mechanism by following the "path to enlightenment" (1), which requires the following: (i) a complete list of components, (ii) the description of all intermediates, (iii) the kinetics of all reactions, and (iv) atomic structures of all components. Some of the gaps in this effort include those listed below.

Have all the components in the DNA packaging reaction been accounted for? Recent description of the role of the T4 late sigma factor, gp55, in DNA packaging (67) suggests that we should be ever vigilant for missing components which may play a vital mechanistic role in DNA packaging. To date, only a single instance of the involvement of a packaging RNA has been described, in ϕ29. The onus is on investigators in the field in general to seek out and remain open to similar new components in all systems under study.

An ongoing controversy deals with the interaction of individual processes, such as prohead maturation and the initiation of DNA translocation, and whether such events are mechanistically coupled. To completely understand any given individual event, we must be aware that processes might depend upon, and in fact be driven by, other seemingly independent processes.

Many of the unknown factors that remain to be investigated relate to the structure and interaction of components. Among these is the question of the order of assembly of packaging ATPase components, DNA substrate, and receptacle prohead. Is the order of assembly in vivo described in λ the same in other systems, with the complex I structure, comprised of terminase and DNA, forming separately from the prohead? Does the prohead play an integral role in DNA maturation, as appears to be the case in P1? Can such interactions help describe the control of DNA processing from concatemers within the crowded environment of the host cell and the mechanism of DNA maturation prior to translocation?

A considerable effort has been directed toward solving the structures of components of the DNA packaging machine at high resolution. Much work remains to be done. The singular example of the solution of the ϕ29 connector structure must (and will) be matched with connectors from other systems. Solution of the structure of packaging ATPase subunits is crucial in building a complete mechanism of their action (24). Atomic resolution of the prohead with embedded portal is as important. Finally, a complete picture of packaging will only be available when structures of all the various intermediates are solved, including that of the intact DNA packaging machine at various points in the ATP hydrolysis cycle.

Broader issues precede such atomic-level concerns. What is the role of symmetry between components in the DNA packaging motor? Does rotation of components play a role, as proposed in several models described above? Will new motifs for molecular motors be revealed as we approach complete elucidation of the mechanism of DNA packaging?

As a relatively mature discipline within molecular biology, research on phage DNA packaging has enormous advantages. The genetics for most systems are well established and the production of large quantities of materials is relatively easy, particularly when compared with eukaryotic systems. This puts the study of dsDNA packaging in the enviable position of being primed for the application of new technologies within the fields of biology and biophysics. Among these are single molecule approaches based upon optical tweezers and atomic force microscopy. Advanced spectroscopy, including fluorescence, RAMAN, EPR, FRET, and many others, seem tailor-made for many of the questions waiting to be answered about the processes involved in packaging. In addition, advances in soft matter physics seem newly capable of providing theoretical insight into the problems and mechanisms involved in DNA packaging, perhaps allowing a return of phage research to its roots in physics.

Lastly, the context in which current and future results are viewed must continually be brought back to their cellular origin. As complex as these processes might seem in vitro, they are perhaps more complex in the in vivo world. This context is crucial if we expect to fully understand these events and processes, and later apply them to other areas of interest.

Acknowledgments

We would like to thank Drs. Michael Feiss, Shelley Grimes, Roger Hendrix, and Theo Odijk for reading a draft of this chapter and for providing helpful comments and suggestions.

References

1. Alberts, B., and R. Miake-Lye. 1992. Unscrambling the puzzle of biological machines: the importance of the details. Cell 68:415–440.

2. Aubrey, K. L., S. R. Casjens, and G. J. Thomas, Jr. 1992. Secondary structure and interactions of the packaged dsDNA genome of bacteriophage P22 investigated by Raman difference spectroscopy. Biochemistry 31:11835–11842.
3. Backhaus, H. 1985. DNA packaging initiation of *Salmonella* bacteriophage P22: determination of cut sites within the DNA sequence coding for gene 3. J. Virol. 55:458–465.
4. Bertani, E., and E. W. Six. 1988. The P2-like phages and their parasite, P4, pp. 73–143. *In* R. Calendar (ed.) The Bacteriophages, vol. 2. Plenum Press, New York.
5. Bijlenga, R. K., D. Scraba, and E. Kellenberger. 1973. Studies on the morphopoiesis of the head of T-even phage. IX. *Tau*-particles: their morphology, kinetics of appearance and possible precursor function. Virology 56:250–267.
6. Bjornsti, M. A., B. E. Reilly, and D. L. Anderson. 1981. In vitro assembly of the *Bacillus subtilis* bacteriophage phi29. Proc. Natl. Acad. Sci. USA 78:5861–5865.
7. Bjornsti, M. A., B. E. Reilly, and D. L. Anderson. 1983. Morphogenesis of bacteriophage phi29 of *Bacillus subtilis*: oriented and quantized in vitro packaging of DNA-protein gp3. J. Virol. 45:383–396.
8. Black, L. W. 1989. DNA packaging in dsDNA bacteriophages. Annu. Rev. Microbiol. 43:267–292.
9. Black, L. W., W. W. Newcomb, J. W. Boring, and J. C. Brown. 1985. Ion etching bacteriophage T4: support for a spiral-fold model of packaged DNA. Proc. Natl. Acad. Sci. USA 82:7960–7964.
10. Black, L. W., and D. J. Silverman. 1978. Model for DNA packaging into bacteriophage T4 heads. J. Virol. 28:643–655.
11. Bowden, D. W., and R. Calendar. 1979. Maturation of bacteriophage P2 DNA in vitro: a complex, site-specific system for DNA cleavage. J. Mol. Biol. 129:1–18.
12. Bowden, D. W., and P. Modrich. 1985. In vitro maturation of circular bacteriophage P2 DNA. Purification of ter components and characterization of the reaction. J. Biol. Chem. 260:6999–7007.
13. Casjens, S., and M. Hayden. 1988. Analysis in vivo of the bacteriophage P22 headful nuclease. J. Mol. Biol. 199:467–474.
14. Casjens, S., W. M. Huang, M. Hayden, and R. Parr. 1987. Initiation of bacteriophage P22 DNA packaging series. Analysis of a mutant that alters the DNA target specificity of the packaging apparatus. J. Mol. Biol. 194:411–422.
15. Casjens, S., E. Wyckoff, M. Hayden, L. Sampson, K. Eppler, S. Randall, E. T. Moreno, and P. Serwer. 1992. Bacteriophage P22 portal protein is part of the gauge that regulates packing density of intravirion DNA. J. Mol. Biol. 224:1055–1074.
16. Catalano, C. E. (ed.) 2003. Viral Genome Packaging. Landes Publishing, Georgetown, Texas.
17. Catalano, C. E., D. Cue, and M. Feiss. 1995. Virus DNA packaging: the strategy used by phage lambda. Mol. Microbiol. 16:1075–1086.
18. Cerritelli, M. E., N. Cheng, A. H. Rosenberg, C. E. McPherson, F. P. Booy, and A. C. Steven. 1997. Encapsidated conformation of bacteriophage T7 DNA. Cell 91:271–280.
19. Chai, S., R. Lurz, and J. C. Alonso. 1995. The small subunit of the terminase enzyme of *Bacillus subtilis* bacteriophage SPP1 forms a specialized nucleoprotein complex with the packaging initiation region. J. Mol. Biol. 252:386–398.
20. Chow, L. T., and A. I. Bukhari. 1978. Heteroduplex electron microscopy of phage Mu mutants containing IS1 insertions and chloramphenicol resistance transposons. Gene 3:333–346.
21. Conway, J. F., R. L. Duda, N. Cheng, R. W. Hendrix, and A. C. Steven. 1995. Proteolytic and conformational control of virus capsid maturation: the bacteriophage HK97 system. J. Mol. Biol. 253:86–99.
22. Cummings, D. J., V. A. Chapman, S. S. DeLong, and N. L. Couse. 1973. Structural aberrations in T-even bacteriophage. 3. Induction of "lollipops" and their partial characterization. Virology 54:245–261.
23. Dannenberg, R., and G. Mosig. 1983. Early intermediates in bacteriophage T4 DNA replication and recombination. J. Virol. 45:813–831.
24. de Beer, T., J. Fang, M. Ortega, Q. Yang, L. Maes, C. Duffy, N. Berton, J. Sippy, M. Overduin, M. Feiss, and C. E. Catalano. 2002. Insights into specific DNA recognition during the assembly of a viral genome packaging machine. Mol. Cell 9:981–991.
25. Dube, P., P. Tavares, R. Lurz, and M. van Heel. 1993. The portal protein of bacteriophage SPP1: a DNA pump with 13-fold symmetry. EMBO J. 12:1303–1309.
26. Duffy, C., and M. Feiss. 2002. The large subunit of bacteriophage lambda's terminase plays a role in DNA translocation and packaging termination. J. Mol. Biol. 316:547–561.
27. Earnshaw, W. C., and S. R. Casjens. 1980. DNA packaging by the double-stranded DNA bacteriophages. Cell 21:319–331.
28. Earnshaw, W. C., J. King, S. C. Harrison, and F. A. Eiserling. 1978. The structural organization of DNA packaged within the heads of T4 wild-type, isometric and giant bacteriophages. Cell 14:559–568.
29. Eickbush, T. H., and E. N. Moudrianakis. 1978. The compaction of DNA helices into either continuous supercoils or folded-fiber rods and toroids. Cell 13:295–306.
30. Eiserling, F. A., E. P. Geiduschek, R. H. Epstein, and E. J. Metter. 1970. Capsid size and deoxyribonucleic acid length: the petite variant of bacteriophage T4. J. Virol. 6:865–876.
31. Feiss, M., R. A. Fisher, M. A. Crayton, and C. Egner. 1977. Packaging of the bacteriophage lambda chromosome: effect of chromosome length. Virology 77:281–293.
32. Feiss, M., S. Frackman, and J. Sippy. 1985. Essential interaction between lambdoid phage 21 terminase and the *Escherichia coli* integrative host factor. J. Mol. Biol. 183:239–246.
33. Franklin, J. L., D. Haseltine, L. Davenport, and G. Mosig. 1998. The largest (70 kDa) product of the bacteriophage T4 DNA terminase gene 17 binds to single-stranded DNA segments and digests them towards junctions with double–stranded DNA. J. Mol. Biol. 277:541–557.
34. Fujisawa, H., and M. Morita. 1997. Phage DNA packaging. Genes Cells 2:537–545.
35. George, M., and A. I. Bukhari. 1981. Heterogeneous host DNA attached to the left end of mature bacteriophage Mu DNA. Nature 292:175–176.

36. Golz, S., and B. Kemper. 1999. Association of Holliday-structure resolving endonuclease VII with gp20 from the packaging machine of phage T4. J. Mol. Biol. 285:1131–1144.
37. Grimes, S., and D. Anderson. 1997. The bacteriophage phi29 packaging proteins supercoil the DNA ends. J. Mol. Biol. 266:901–914.
38. Grimes, S., P. J. Jardine, and D. Anderson. 2002. Bacteriophage phi29 DNA packaging. Adv. Virus Res. 58:255–294.
39. Guasch, A., J. Pous, B. Ibarra, F. X. Gomis-Ruth, J. M. Valpuesta, N. Sousa, J. L. Carrascosa, and M. Coll. 2002. Detailed architecture of a DNA translocating machine: the high-resolution structure of the bacteriophage phi29 connector particle. J. Mol. Biol. 315:663–676.
40. Guo, P., S. Grimes, and D. Anderson. 1986. A defined system for in vitro packaging of DNA-gp3 of the *Bacillus subtilis* bacteriophage phi29. Proc. Natl. Acad. Sci. USA 83:3505–3509.
41. Hamada, K., H. Fujisawa, and T. Minagawa. 1986. A defined in vitro system for packaging of bacteriophage T3 DNA. Virology 151:119–123.
42. Hang, J. Q., B. F. Tack, and M. Feiss. 2000. ATPase center of bacteriophage lambda terminase involved in post-cleavage stages of DNA packaging: identification of ATP–interactive amino acids. J. Mol. Biol. 302:777–795.
43. Harshey, R. M. 1988. Phage Mu, pp. 193–234. *In* R. Calendar (ed.) The Bacteriophages, vol. 1. Plenum Press, New York.
44. Hendrix, R. W. 1978. Symmetry mismatch and DNA packaging in large bacteriophages. Proc. Natl. Acad. Sci. USA 75:4779–4783.
45. Hendrix, R. W. 1985. Shape determination in virus assembly: the bacteriophage example, pp. 169–203. *In* S. Casjens (ed.) Virus Structure and Assembly, vol. 1. Jones and Bartlett, Boston, Mass.
46. Hendrix, R.W., and S. R. Casjens. 1975. Assembly of bacteriophage lambda heads: protein processing and its genetic control in petite lambda assembly. J. Mol. Biol. 91:187–199.
47. Hohn, T. 1976. Packaging of genomes in bacteriophages: a comparison of ssRNA bacteriophages and dsDNA bacteriophages. Phil. Trans. R. Soc. Lond. B Biol. Sci. 276:143–150.
48. Hud, N. V. 1995. Double-stranded DNA organization in bacteriophage heads: an alternative toroid-based model. Biophys. J. 69:1355–1362.
49. Hud, N. V., and K. H. Downing. 2001. Cryoelectron microscopy of lambda phage DNA condensates in vitreous ice: the fine structure of DNA toroids. Proc. Natl. Acad. Sci. USA 98:14925–14930.
50. Hwang, Y., and M. Feiss. 1995. A defined system for in vitro lambda DNA packaging. Virology 211:367–376.
51. Jardine, P. J., and D. H. Coombs. 1998. Capsid expansion follows the initiation of DNA packaging in bacteriophage T4. J. Mol. Biol. 284:661–672.
52. Jardine, P. J., M. C. McCormick, C. Lutze-Wallace, and D. H. Coombs. 1998. The bacteriophage T4 DNA packaging apparatus targets the unexpanded prohead. J. Mol. Biol. 284:647–659.
53. Kellenberger, E. 1980. Control mechanisms in the morphogeneses of bacteriophage heads. Biosystems 12:201–223.
54. Kellenberger, E., E. Carlemalm, J. Sechaud, and A. Ryter. 1986. Considerations on the condensation and the degree of compactness in non-eukaryotic DNA–containing plasmids, pp. 11–25. *In* C. O. Gualerzi and C. L. Pon (eds.) Bacterial Chromatin, vol. 1. Springer, Berlin.
55. Keppel, F., O. Fayet, and C. Georgopoulos. 1988. Strategies of bacteriophage replication, pp. 145–262. *In* R. Calendar (ed.) The Bacteriophages, vol. 2. Plenum Press, New York.
56. Kindt, J., S. Tzlil, A. Ben-Shaul, and W. M. Gelbart. 2001. DNA packaging and ejection forces in bacteriophage. Proc. Natl. Acad. Sci. USA 98:13671–13674.
57. King, J., and S. Casjens. 1974. Catalytic head assembling protein in virus morphogenesis. Nature 251:112–119.
58. Kinosita, K., Jr., R. Yasuda, H. Noji, S. Ishiwata, and M. Yoshida. 1998. F1-ATPase: a rotary motor made of a single molecule. Cell 93:21–24.
59. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685.
60. Laemmli, U. K., and M. Favre. 1973. Maturation of the head of bacteriophage T4. I. DNA packaging events. J. Mol. Biol. 80:575–599.
61. Lepault, J., J. Dubochet, W. Baschong, and E. Kellenberger. 1987. Organization of double-stranded DNA in bacteriophages: a study by cryo-electron microscopy of vitrified samples. EMBO J. 6:1507–1512.
62. Lin, H., V. B. Rao, and L. W. Black. 1999. Analysis of capsid portal protein and terminase functional domains: interaction sites required for DNA packaging in bacteriophage T4. J. Mol. Biol. 289:249–260.
63. Luftig, R. B., and C. Ganz. 1972. Bacteriophage T4 head morphogenesis. IV. Comparison of gene 16-, 17-, and 49-defective head structures. J. Virol. 10:545–554.
64. Luftig, R. B., and N. P. Lundh. 1973. Bacteriophage T4 head morphogenesis. Isolation, partial characterization, and fate of gene 21-defective *tau*-particles. Proc. Natl. Acad. Sci. USA 70:1636–1640.
65. Luftig, R. B., W. B. Wood, and R. Okinaka. 1971. Bacteriophage T4 head morphogenesis. On the nature of gene 49-defective heads and their role as intermediates. J. Mol. Biol. 57:555–573.
66. Lurz, R., E.V. Orlova, D. Gunther, P. Dube, A. Droge, F. Weise, M. van Heel, and P. Tavares. 2001. Structural organisation of the head-to-tail interface of a bacterial virus. J. Mol. Biol. 310:1027–1037.
67. Malys, N., D. Y. Chang, R. G. Baumann, D. Xie, and L. W. Black. 2002. A bipartite bacteriophage T4 SOC and HOC randomized peptide display library: detection and analysis of phage T4 terminase (gp17) and late sigma factor (gp55) interaction. J. Mol. Biol. 319:289–304.
68. Mellado, R. P., M. A. Penalva, M. R. Inciarte, and M. Salas. 1980. The protein covalently linked to the 5′ termini of the DNA of *Bacillus subtilis* phage phi29 is involved in the initiation of DNA replication. Virology 104:84–96.
69. Mellado, R. P., and M. Salas. 1983. Initiation of phage phi29 DNA replication by the terminal protein modified at the carboxyl end. Nucleic Acids Res 11:7397–7407.

70. Mitchell, M. S., S. Matsuzaki, S. Imai, and V. B. Rao. 2002. Sequence analysis of bacteriophage T4 DNA packaging/terminase genes 16 and 17 reveals a common ATPase center in the large subunit of viral terminases. Nucleic Acids Res. 30:4009–4021.
71. Mizuuchi, K., B. Kemper, J. Hays, and R. A. Weisberg. 1982. T4 endonuclease VII cleaves Holliday structures. Cell 29:357–365.
72. Mullaney, J. M., and L. W. Black. 1998. Activity of foreign proteins targeted within the bacteriophage T4 head and prohead: implications for packaged DNA structure. J. Mol. Biol. 283:913–929.
73. Muller, D. J., A. Engel, J. L. Carrascosa, and M. Velez. 1997. The bacteriophage phi29 head–tail connector imaged at high resolution with the atomic force microscope in buffer solution. EMBO J. 16:2547–2553.
74. Odijk, T. 1998. Hexagonally packed DNA within bacteriophage T7 stabilized by curvature stress. Biophys. J. 75:1223–1227.
75. Oster, G., and H. Wang. 2003. Rotary protein motors. Trends Cell Biol. 13:114–121.
76. Overman, S. A., K. L. Aubrey, K. E. Reilly, O. Osman, S. J. Hayes, P. Serwer, and G. J. Thomas, Jr. 1998. Conformation and interactions of the packaged double-stranded DNA genome of bacteriophage T7. Biospectroscopy 4:S47–56.
77. Pruss, G. J., and R. Calendar. 1978. Maturation of bacteriophage P2 DNA. Virology 86:454–467.
78. Raman, C. S., S. J. Hayes, B. T. Nall, and P. Serwer. 1993. Energy stored in the packaged DNA of bacteriophage T7. Biophys. J. (Suppl.) 64:A12.
79. Rao, V. B., and L. W. Black. 1985. DNA packaging of bacteriophage T4 proheads in vitro. Evidence that prohead expansion is not coupled to DNA packaging. J. Mol. Biol. 185:565–578.
80. Rekosh, D. M., W. C. Russell, A. J. Bellet, and A. J. Robinson. 1977. Identification of a protein linked to the ends of adenovirus DNA. Cell 11:283–295.
81. Roeder, G. S., and P. D. Sadowski. 1977. Bacteriophage T7 morphogenesis: phage-related particles in cells infected with wild-type and mutant T7 phage. Virology 76:263–285.
82. Schmutz, M., D. Durand, A. Debin, Y. Palvadeau, A. Etienne, and A. R. Thierry. 1999. DNA packing in stable lipid complexes designed for gene transfer imitates DNA compaction in bacteriophage. Proc. Natl. Acad. Sci. USA 96:12293–12298.
83. Serwer, P. 1975. Buoyant density sedimentation of macromolecules in sodium iothalamate density gradients. J. Mol. Biol. 92:433–448.
84. Serwer, P. 1986. Arrangement of double-stranded DNA packaged in bacteriophage capsids. An alternative model. J. Mol. Biol. 190:509–512.
85. Serwer, P. 1988. The source of energy for bacteriophage DNA packaging: an osmotic pump explains the data. Biopolymers 27:165–169.
86. Shibata, H., H. Fujisawa, and T. Minagawa. 1987. Characterization of the bacteriophage T3 DNA packaging reaction in vitro in a defined system. J. Mol. Biol. 196:845–851.
87. Simpson, A. A., Y. Tao, P. G. Leiman, M. O. Badasso, Y. He, P. J. Jardine, N. H. Olson, M. C. Morais, S. Grimes, D. L. Anderson, T. S. Baker, and M. G. Rossmann. 2000. Structure of the bacteriophage phi29 DNA packaging motor. Nature 408:745–750.
88. Smith, D. E., S. J. Tans, S. B. Smith, S. Grimes, D. L. Anderson, and C. Bustamante. 2001. The bacteriophage phi29 portal motor can package DNA against a large internal force. Nature 413:748–752.
89. Son, M., S. J. Hayes, and P. Serwer. 1989. Optimization of the in vitro packaging efficiency of bacteriophage T7 DNA: effects of neutral polymers. Gene 82:321–325.
90. Steven, A. C., E. Couture, U. Aebi, and M. K. Showe. 1976. Structure of T4 polyheads. II. A pathway of polyhead transformation as a model for T4 capsid maturation. J. Mol. Biol. 106:187–221.
91. Steven, A. C., H. L. Greenstone, F. P. Booy, L. W. Black, and P. D. Ross. 1992. Conformational changes of a viral capsid protein. Thermodynamic rationale for proteolytic regulation of bacteriophage T4 capsid expansion, cooperativity, and super-stabilization by soc binding. J. Mol. Biol. 228:870–884.
92. Stroud, R. M., P. Serwer, and M. J. Ross. 1981. Assembly of bacteriophage T7. Dimensions of the bacteriophage and its capsids. Biophys. J. 36:743–757.
93. Tao, Y., N. H. Olson, W. Xu, D. L. Anderson, M. G. Rossmann, and T. S. Baker. 1998. Assembly of a tailed bacterial virus and its genome release studied in three dimensions. Cell 95:431–437.
94. Tavares, P., M. A. Santos, R. Lurz, G. Morelli, H. de Lencastre, and T. A. Trautner. 1992. Identification of a gene in *Bacillus subtilis* bacteriophage SPP1 determining the amount of packaged DNA. J. Mol. Biol. 225:81–92.
95. Turnquist, S., M. Simon, E. Egelman, and D. Anderson. 1992. Supercoiled DNA wraps around the bacteriophage phi29 head–tail connector. Proc. Natl. Acad. Sci. USA 89:10479–10483.
96. Valpuesta, J. M., and J. L. Carrascosa. 1994. Structure of viral connectors and their function in bacteriophage assembly and DNA packaging. Q. Rev. Biophys. 27:107–155.
97. Watson, J. D. 1972. Origin of concatemeric T7 DNA. Nat. New. Biol. 239:197–201.
98. Widom, J., and R. L. Baldwin. 1980. Cation-induced toroidal condensation of DNA studies with $Co^{3+}(NH_3)_6$. J. Mol. Biol. 144:431–453.
99. Widom, J., and R. L. Baldwin. 1983. Tests of spool models for DNA packaging in phage lambda. J. Mol. Biol. 171:419–437.
100. Wieczorek, D. J., and M. Feiss. 2001. Defining cosQ, the site required for termination of bacteriophage lambda DNA packaging. Genetics 158:495–506.
101. Yap, N. L., and V. B. Rao. 1996. Novel mutants in the 5′ upstream region of the portal protein gene 20 overcome a gp40-dependent prohead assembly block in bacteriophage T4. J. Mol. Biol. 263:539–550.

7

General Aspects of Lysogeny

ALLAN CAMPBELL

Cell lines harboring latent viruses are common both in eukaryotes and in prokaryotes. Prokaryotes harboring latent phages are *lysogenic*, and the latent form of the phage is called a *prophage*. Phages that can enter such a latent state are called *temperate*, although some authors erroneously refer to them as *lysogenic*.

Lysogeny was discovered because some bacterial isolates spontaneously produce small amounts of infectious phage. It was later shown that lysogenic bacteria could arise during laboratory infection.

Lysogeny is heritable within a bacterial lineage. When a lysogenic cell divides, both daughter cells harbor the prophage. Very occasionally, a lysogenic cell spontaneously lyses and liberates phage. Lwoff and Gutmann (5) demonstrated this by separating bacterial cells at division by micromanipulation and testing their descendants. Their results eliminated the possibility that lysogeny might be an artifact of reinfection within a culture.

The existence of lysogeny poses three basic questions, which have diverse answers among the various known temperate phages: (i) What is the physical nature of the prophage? (ii) What ensures that the prophage replicates during each division cycle and is segregated to both daughters at cell division? (iii) Since the prophage must contain all the genes needed to carry out a lytic cycle, what restrains it from doing so? The first two questions are very closely connected. The third was a cornerstone of Jacob and Monod's (4) proposal that transcription initiation is a general mechanism for controlling gene expression.

Although all temperate phages must achieve regular inheritance and control of lytic functions, they accomplish this in diverse ways. The purpose here is neither to chronicle all this variety nor to detail the mechanisms of particular phages discussed in other chapters. This chapter attempts rather to outline some frequently used mechanisms, emphasizing their relevance to the general goals for which they were selected.

Nature and Mode of Replication of the Prophage

Nature of the Prophage

In lysogens of most temperate phages (including coliphages λ and Mu-1), the complete phage genome is inserted into the continuity of the bacterial chromosome. Prophage DNA is replicated and segregated as part of the chromosome. No phage-specified enzymes are needed for prophage replication. Synthesis of the replication enzymes used during lytic infection is repressed, and their untimely synthesis can be deleterious to the cell.

The major alternative mode of lysogeny (used by phage P1) is for the phage genome to become established as a plasmid, separate from the chromosome, which segregates in a regular manner at cell division. The plasmid form is typically circular and encodes some of the enzymes needed for its own replication. In phage P1, both the replication genes and the replication origin used by the prophage are different from those used in the lytic cycle.

Table 7-1 lists some common phages classified according to the state of the prophage. Among the inserted prophages, two very distinct mechanisms of insertion are recognized. Three groups of phages (exemplified by λ, Mu-1, and P1, respectively) will now be discussed in more depth.

Prophages Inserted by Reciprocal, Site-Specific Recombination

λ is one of the many phages that insert their DNA into the host chromosome by reciprocal recombination between the chromosome and a circular form of the phage DNA. The overall reaction comprises cointegration of the two circles. The recombination takes place between a unique site on the phage DNA and a unique site on the *E. coli*

Table 7-1 Modes of Prophage Maintenance

Mode	Example
Insertion	
By site-specific recombination	Coliphage λ *Salmonella* phage P22 Mycobacterial phage L5 Vibrio phage CTX Phi *Streptomyces* phage ϕC31
By transposition	Mu-1 D108
Plasmid formation	Coliphages P1, P7 *Salmonella* phage D6

chromosome (figure 7-1). Host and phage DNA are identical in sequence for 15 bp at the crossover point. Fifteen base pairs is not enough to allow recombination by the general recombinase of the host (*rec*) or of the phage (*red*). Phage λ encodes a protein (integrase) that efficiently mediates reciprocal recombination at these sites. Integrase is made in greatest abundance in cells that are recovering from infection. After lysogeny is established, the integrase gene is turned off.

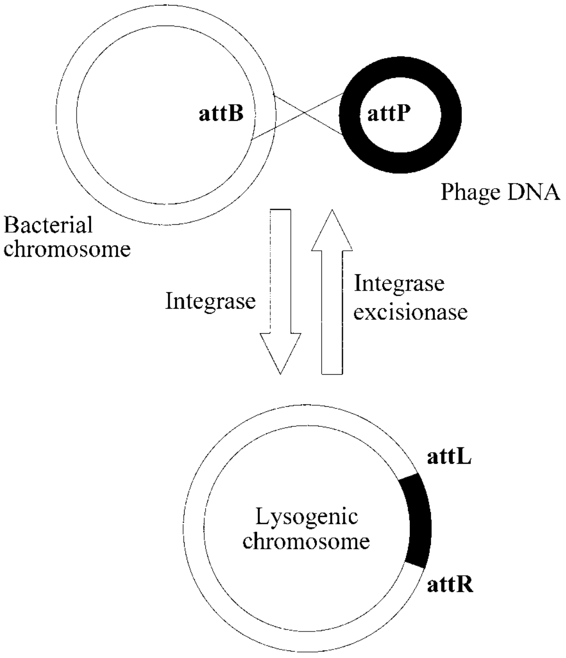


Figure 7-1 Integration by site-specific recombination. Integration occurs between the *attP* site of the phage and the *attB* site of the host. The two circular molecules form a cointegrate, with junction points *attL* (left prophage end) and *attR* (right prophage end). Integration is carried out by a phage-coded protein, integrase. In many phages (including λ), the reverse reaction (excision) frequently requires a second phage-coded protein, excisonase. Host proteins contributing to the reaction are not shown.

This is precisely what is expected of a mechanism to promote lysogeny: it should go on after infection, so that every surviving cell has an inserted prophage. It is unnecessary in an established lysogen, where insertion has already occurred. Finally, if lysogeny is disrupted and the cell switches back to the lytic cycle (as occasionally happens spontaneously and can be induced), the prophage should be excised from the chromosome. For reasons that are not completely understood at a biochemical level, the insertion reaction is not directly reversible. To excise the prophage, another phage protein, excisionase, is needed. Following induction, the two proteins are produced coordinately.

The integrase/excisionase system puts the phage in charge of the timing and direction of site-specific recombination. If the phage and bacterium had a longer stretch of homology and depended on general recombinases, insertion and excision would happen rarely and haphazardly.

The effective irreversibility of the insertion reaction in the absence of excisionase has suggested that it might be adapted to insert DNA into specific sites of the human genome for gene therapy (7). The idea is to promote stable insertion into some human sequence resembling the bacterial site. The integration system of *Streptomyces* phage ϕC31 has been used for this purpose.

λ inserts in intergenic DNA, but many other temperate phages, including some natural relatives of λ, insert at sites that are within structural genes (21) or tRNA genes (P22). What these sites have in common is an interrupted (frequently imperfect) dyad symmetry centered on the crossover point. In tRNA genes, this configuration is found in the anticodon loop, which is used not only by the λ-related phage P22 but also by coliphage 186, *Haemophilus* phage HP1 and mycobacterium phage L5.

The λ integrase reaction proceeds through two successive strand exchanges placed 7 bp apart. The symmetry of the DNA site allows equivalent recognition at the two exchange points. The sites used by some phages have no obvious symmetry. The satellite coliphages P4 and its relatives insert into tRNA genes, but in the TψC loop near the 3′ end, with no DNA symmetry centered on the insertion point. P4 uses an integrase of the same superfamily as λ integrase, but the two integrases are barely related. *Streptomyces* phage ϕC31 inserts into a site with no apparent symmetry, using a site-specific recombinase from a different superfamily. Thus the site-specific recombination systems used for insertion/excision employ various mechanisms, but all produce the same final result.

Insertion by site-specific recombinases requires a double-stranded circle of phage DNA (figure 7-1). As packaged in the virion, the DNAs of most temperate phages are not double-stranded circles. They are generally converted to that form early in phage development. λ, for example, has linear DNA with projecting complementary single-stranded 5′ ends, which pair following infection, allowing ligation to a double-stranded circle. The cholera phage CTX Phi has

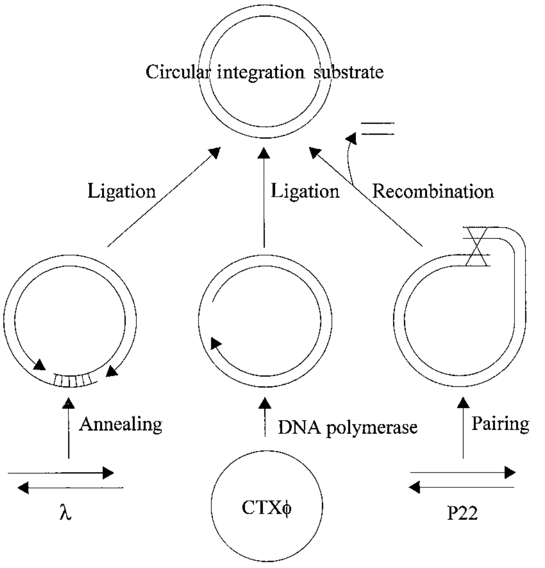


Figure 7-2 Some pathways to the double-stranded, circular integration substrate. See text for details.

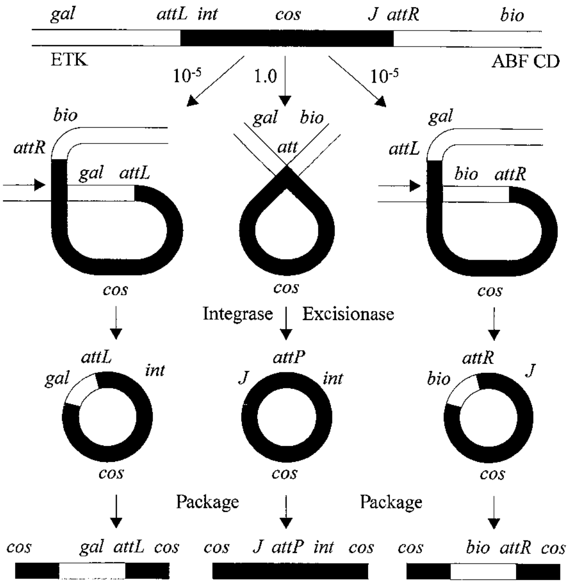


Figure 7.3 Formation of specialized transducing phages by bacteriophage λ. The central column shows normal excision of the phage genome, mediated by integrase and excisionase. For every such excision, there are about 10^{-5} abnormal excisions, where either *gal* or *bio* is effectively cloned into λ. These are rare because they require breakage and joining of heterologous DNA. The reciprocal product (a deleted chromosome), which may or may not be formed in the same event, is not shown. The presence of the *cos* site allows the excised DNA to be packaged into virions and injected into other cells. Since λ is cut at *cos* during packaging, part of *cos* occurs at each end of the packaged DNA.

single-stranded circles in the virion, but like most phages, it replicates as a double-stranded circle. Phage P22 has headful packaging, where the DNA molecule injected from the virion has a direct repetition (different among virions) of double-stranded DNA at the ends. Within the cell, homology-dependent recombination produces a circle (the same for all virions). Some of these pathways to the circular integration substrate are illustrated in figure 7-2. In all these cases, the circular form is an intermediate in lytic development as well as in integration.

Although site-specific recombinases of the integrase family occur in eukaryotes, they are not known to be used in viral integration. The dependent parvoviruses (so-called adeno-associated viruses) insert with high preference in a specific small human chromosomal segment, but there is no indication that the mechanism has much in common with phage insertion (8).

Specialized Transduction

Transduction is the transfer of genes from one host cell to another by a virus. There are two types: generalized (which results from errors in packaging) and specialized (which results from errors in excision from the chromosome of a lysogenic bacterium). Generalized transduction is not directly related to lysogeny. It is most easily observed with temperate phages, but only because the potential transductants formed by a virulent phage are generally destroyed by lysis. This lysis is not caused by the transducing particles, which usually contain only host DNA, but rather by the phage particles in the lysate.

Specialized transduction occurs only with temperate phages, and is observed only in lysates produced by inducing lytic development in a lysogen, not in lysates produced by infection. Specialized transduction results from rare faulty excision from the lysogenic chromosome. Instead of the precise excision mediated by integrase and excisionase (figure 7-1), an unknown enzymology causes breakage and joining in heterologous DNA to produce packageable chimeric genomes, part phage and part host. This can be viewed as a natural form of gene cloning. Whereas any part of the host genome can be transferred in generalized transduction, specialized transduction is possible only for a limited portion of the host chromosome flanking the phage insertion site.

Figure 7-3 shows specialized transduction by λ of some nearby genes—*gal* from one side and *bio* from the other. λ *gal* and λ *bio* arise in separate events. Several constraints on the process may be noted: (i) The genome of a specialized transducing phage is a connected segment of the lysogenic chromosome. Therefore, any λ carrying *gal* will also carry all the genes between *gal* and *att*λ, and any λ carrying the distal *galE* gene must also carry *galK*. (ii) The amount of DNA is restricted by the packaging limits of the phage. λ can package up to 110% of its 48.7 kb genome, so no gene can be packaged that is farther than that from λ *cos*. By extension, if a λ has accidentally inserted somewhere other than its normal site, it will now transduce only genes close to the new site. (iii) Specialized transducing phages are frequently missing some phage genes. As implied by figure 7-3, acquisition of host DNA to the left of the prophage is accompanied by loss of genes from the right end of the prophage. The phage genes that are eliminated may be vital for the viral life cycle, so propagation of the specialized transducing variant can take place only in cells coinfected with a complete λ. (iv) The packaging site, *cos*, cannot be lost because it is needed in cis. Therefore, all *gal*-transducers will include genes *int-cos*, and all *bio* transducers will include *cos-J*.

The last line of figure 7-3 compares the structures of λ *gal* and λ *bio* DNA with that of normal λ, as packaged in the virion. The termini (produced by cutting at *cos*) are the same. λ *gal*s have a segment of host DNA whose left margin may vary widely but whose right end is always at the phage insertion site; in λ *bio*s, the host DNA runs rightward from the insertion site.

Figure 7-4 shows insertion by the λ *gal* specialized transducing phage. λ *gal* may insert into the chromosome either by general, homology-dependent recombination or by site-specific recombination at *att*. The insertion is sometimes accompanied by insertion of a coinfecting normal λ (not shown in figure 7-4). The requirements for the integrase reaction using the hybrid site *attL* on the λ *gal* are different from those for a phage *attP* site; this generates a helping effect by the coinfecting λ. The product of lysogenization contains two copies of *gal*: the indigenous copy already present before infection (shown as *gal*$^-$ in figure 7-4) and the added copy from the λ *gal* (*gal*$^+$ in figure 7-4). Homology-dependent recombination within the chromosome can "loop out" one copy or the other, so that lysogens of λ *gal* are inherently unstable.

Insertion and excision by λ *gal* have served as a model for replacement of genes by cloned, mutated copies. Whereas λ insertion requires integrases, λ *gal* insertion can also take place through homologous recombination with bacterial DNA. The haploid segregants can be either Gal$^-$ (reversal) or Gal$^+$ (replacement) depending on which side of *gal* the crossover occurs to loop out the λ *gal*.

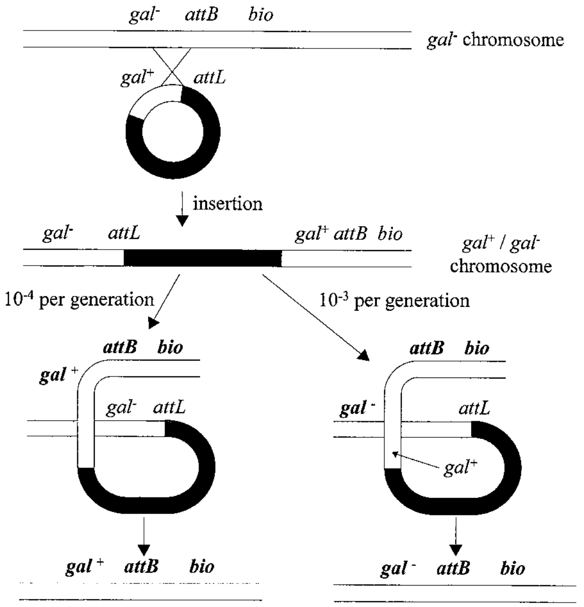


Figure 7-4 Insertion and looping out of specialized transducing phage. The λ *gal*$^+$ can insert into the chromosome either by site-specific recombination at *att* or by homology-dependent recombination in host DNA between *gal* and *att*λ. The product (a heterogenote) has two copies of *gal*, separated by λ DNA. During growth of the strain, the λ *gal* can occasionally loop out through homology-dependent recombination. Depending on the position of the crossover point, the haploid derivative can carry either the *gal* gene of the original chromosome or that brought in by the phage. Looping out to give the chromosomal gene happens at about 10^{-3} per generation. The rate for the phage-derived gene (in this case, *gal*$^+$) is lower and depends on the extent of homology to the left of *gal*. The reciprocal product (an excised λ *gal* which, if formed, does not replicate and is therefore diluted out by growth) is not shown.

Prophage Insertion by Transposition

The most familiar example of insertion by transposition is coliphage Mu-1. There are fundamental differences between the life cycles of Mu-1 and λ. In Mu-1, there is no circular form of the viral DNA. Also, viral DNA is always inserted into host DNA, even in the virion. Both replication and insertion result from transposition of viral DNA from one host site to another.

Transposition is diagrammed in figure 7-5. Transposition has two possible outcomes, replicative and nonreplicative, of which only the nonreplicative concerns us here. In cells destined to lyse, the replicative mode is used and reused, moving viral DNA into different sites and rearranging the chromosomal DNA many times before the cell finally lyses and liberates a crop of progeny virions. During packaging, DNA is cut at nonspecific positions distal to the termini of viral DNA, so that each virion contains a small segment of host DNA (typically about 2 kb) with the 39 kb viral genome inserted into it. In cells destined for lysogeny, nonreplicative transposition splices viral DNA into the chromosome. Any site on the chromosome can be used as a target for transposition, so each lysogen has a unique insertion site. Insertions within genes generally disrupt gene function, which constitutes a kind of mutation, hence the name "mutator phage."

Mu-1 does not produce specialized transducing phages. Every virion contains about 2 kb of flanking DNA; but these are separated from viral DNA in the first cycle of transposition following infection. However, Mu-1 has been modified by genetic engineering to allow in vivo cloning of host DNA segments of 5 kb or more (figure 7-6).

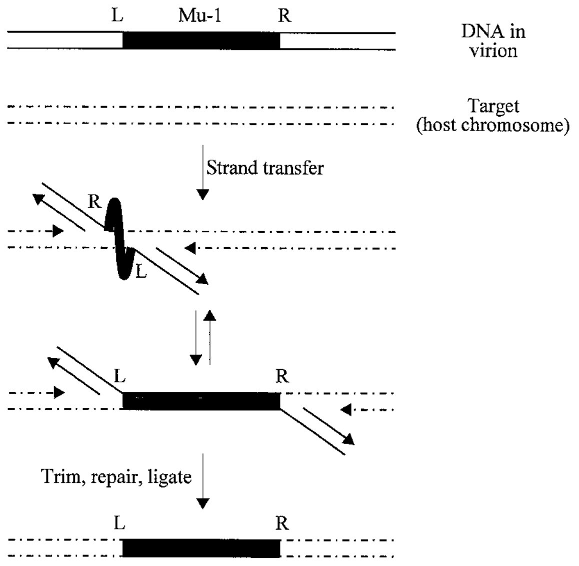


Figure 7-5 Chromosomal insertion of phage Mu-1 DNA by a transposition mechanism. In the virion, Mu-1 DNA is attached to host DNA at both ends: less than 100 bp on the left, and about 2 kb on the right. Any part of the chromosome can be used as target. The first step of transposition is strand transfer, where the 3′ ends of Mu are ligated to opposite strands of target DNA 5 bp apart, leaving free 3′ ends of the chromosomal DNA and free 5′ ends of the host DNA from the virion. The strand transfer product is redrawn in the next line for ease of visualization. Repair synthesis initiated from the 3′ ends of target DNA fill the gap, the host DNA that entered from the virion is trimmed off, and the target DNA 3′ ends are ligated to the 5′ Mu-1 ends. Mu-1 can also undergo replicative transposition where the donor DNA is not trimmed and the transposon is replicated; however, this is only fruitful with a circular donor and does not occur in the initial transposition from linear virion DNA.

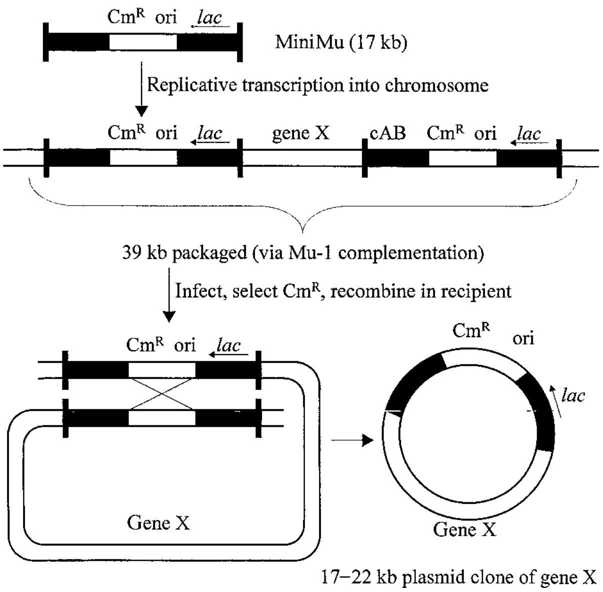

Cm^R ori lac
MiniMu (17 kb)
Replicative transcription into chromosome
Cm^R ori lac
gene X
cAB
Cm^R ori lac
39 kb packaged (via Mu-1 complementation)
Infect, select Cm^R, recombine in recipient
Cm^R ori lac
Cm^R
ori
lac
Gene X
Gene X
17–22 kb plasmid clone of gene X

Figure 7-6 In vivo cloning with mini Mu *lac*. The initial mini Mu *lac* includes, between Mu-1 ends, a selectable marker (Cm^R), a plasmid replication origin (*ori*), and a *lacZ* gene suitable for constructing translational fusion. It can be packaged by complementation and introduced into a cell harboring a complete Mu-1 prophage. Induction of phage development leads to repeated rounds of replicative transposition (chapter 30) of both Mu-1 and mini Mu *lac*, so that the cell population will include cells where any desired gene is flanked by two directly oriented mini Mu *lac* prophages at an appropriate distance so that the two mini Mu *lacs* plus intervening DNA can be packaged into a Mu-1 virion. In the recipient, homology-dependent recombination excises a plasmid containing the cloned gene X. Based on Groisman et al. (2). The method has also been modified to use a mini Mu with λ *cos* sites, packageable by a complementing λ, which allows a cloning of longer segments (1).

Among eukaryotic viruses, the retroviruses employ an insertion mechanism almost identical to that of Mu-1. The rest of their life cycle is quite different, however.

Lysogenization by Plasmid Formation

Phage P1 is the prototype of lysogenization by plasmid formation. Like P22, P1 DNA is packaged as circularly permuted, terminally repetitious molecules, which circularize soon after infection through recombination between the terminal segments (cf. figure 7.2). In cells destined for lysogeny, the circular molecules persist as plasmids, replicating once during each cell cycle and segregating regularly at cell division, like the bacterial chromosome. The replication origin for plasmid replication (which is distant from the origin used in lytic development) is closely linked to the determinants for orderly segregation.

P1 encodes a site-specific recombinase, Cre, that acts on a unique, completely symmetrical site (*lox*) on P1 DNA. There are no *lox* sites on the bacterial chromosome, and the Cre *lox* system is not used for insertion (except for rare, aberrant insertions into pseudolox sites). Cre serves two functions in the phage life cycle: (a) Cre can mediate the circularization of viral DNA that happens early after infection (cf. figure 7-2). That only occurs among those molecules where the *lox* site is within the terminal repetition, but because P1 initiates packaging at a preferred site (*pac*), such molecules are frequent. This circularization is needed for lytic as well as lysogenic development. (b) Cre facilitates segregation of the plasmid at cell division, thereby increasing the stability of the lysogenic state. During or after P1 replication, homologous recombination between the two daughter molecules may create a dimer. At cell division, the dimer can only go to one of the two daughter cells, leaving the other plasmidless. Site-specific recombination at *lox* resolves the dimer into two monomers, which can then segregate normally. The rate of spontaneous loss of the P1 plasmid is less than 10^{-5} per cell division with wild-type P1 and 10^{-2} when the *lox* has been deleted (6).

Some eukaryotic DNA viruses, including Epstein–Barr and papilloma, can also establish themselves as plasmids that replicate within the nucleus.

Defective Phages and Prophages

Natural bacterial strains frequently harbor incomplete prophages (also called *defective* or *cryptic*) as well as complete ones. In the wild, bacteria are continually subject to infection, reinfection, and lysogenization. As lysogenic bacteria grow, their prophages experience deletions, mutations, or insertions within the prophage genome. There is generally no selection against such events, so they accumulate over evolutionary time and are frequently observed in laboratory lysogens as well. Defective lysogens were recognized early in the modern era of phage biology as strains that retained some properties of lysogens (such as immunity to infection by phage of the carried type) but did not liberate plaque-forming particles. Whole genome sequencing has underscored their frequency in natural populations. For example, the K-12 strain of *E. coli* harbors (in addition to λ), four or five λ-related defective prophages, and the enterohemorrhagic *E. coli* H7:0157 has seven (including a defective prophage at the λ site; 3). These strains also carry several elements related to phages P4 and P2. Several percent of the typical bacterial genome is phage-derived. A similar fraction of the mammalian genome is virus- (especially retrovirus-) related.

Some of these elements have attracted special attention because they contain, between *attL* and *attR* sites, both phage-related integrase genes and genes with special bacterial functions, such as pathogenicity. It is generally not known how the specialized genes became associated with the element or whether the complete element inserts or excises.

Gene Regulation by Repression

Regulation in a Lysogen

In a lysogenic bacterium, the genes of the phage lytic cycle must be turned off, or the cell would die. Turnoff is caused by a phage-coded repressor (usually a protein, but in some cases, as with phage P4, an antisense RNA). The repressor prevents transcriptional initiation from key phage promoters. This turns off almost all phage transcription. Repressor acts directly on only a few promoters (in λ, there are two). These are the promoters from which lytic development is initiated. Phage genes expressed late in lytic development are turned off indirectly, by the lack of early gene products needed to activate their expression. The skeletal control system in λ (where the repressor is coded by the *cI* gene) is shown in figure 7-7. Repressor not only turns off transcription from the lytic promoters pL and pR but also promotes its own transcription. This is accomplished by specific binding to a cluster of three repressor-binding sites at each of the two operators. Late genes determining cell lysis and virion components are not expressed in the lysogen because their transcription is turned on only when the Q protein antiterminates the pR′ transcript, and the *Q* gene is transcribed from pR.

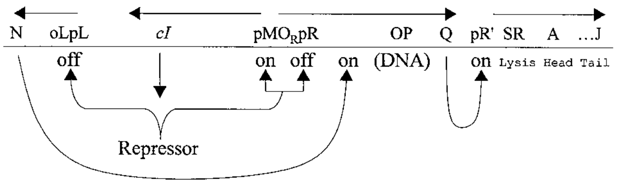


Figure 7-7 Outline of gene control in a λ lysogen. Repressor (product of gene *cI*) binds to operator sites (oL, oR, also known as O_L and O_R, respectively), turning off transcription from pL (also known as P_L) and pR (also known as P_R) and turning on *cI* transcription from pM. Repressor blocks transcription of late genes indirectly, by preventing transcription from gene *Q*, whose product turns on late gene transcription of gene *N*, whose product antiterminates early gene transcription.

The lysogenic state is quite stable. About once in 10^4 cell divisions, lytic development is spontaneously activated, and the cell lyses and liberates phage. In many phages (including λ), lytic development can be induced experimentally. The most convenient method of induction is to heat a strain that harbors a prophage with a *ts* mutation in the *cI* gene, so that the repressor is thermosensitive and denatures at high temperature. A more natural method is to treat the cell with various agents (such as ultraviolet light) that activate the bacterial SOS system.

The SOS system is a global regulatory system responsive to DNA damage. The basic circuitry is shown in figure 7-8. When DNA is damaged by irradiation or other means, single-strand oligonucleotides are formed as breakdown products. These activate a coprotease activity of the RecA protein (which also mediates homology-dependent recombination). This coprotease activity promotes cleavage of a master control protein, LexA. RecA is considered a coprotease rather than a protease because the LexA protein can undergo spontaneous cleavage (autodigestion) under nonphysiological conditions, such as extreme pH, in the absence of RecA. LexA represses, directly or indirectly, a large number of genes whose products facilitate repair and recovery from DNA damage. So when LexA is cleaved in response to DNA damage, these proteins are synthesized. Many phage repressors (including λ's) mimic LexA in their susceptibility to coproteolysis by activated RecA. When repressor is thus inactivated, lytic development ensues in almost every cell of a treated culture. Spontaneous phage production is RecA-dependent and therefore presumed to result from rare sporadic DNA damage, such as may occur when a replication fork occasionally stalls.

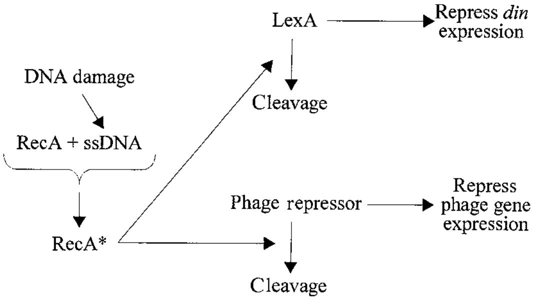


Figure 7-8 The bacterial SOS system and its relation to induction of lytic development. DNA damage causes the disintegration of some DNA to smaller degradation products, of which single-stranded DNA is especially effective. Single-stranded DNA activates the RecA protein to a form RecA* whose coprotease activity accelerates the spontaneous cleavage of the LexA protein into two inactive polypeptides. The normal function of LexA is to repress (either directly or indirectly) the transcription of damage-inducible (*din*) genes of the host, many of which function in DNA repair. The repressor proteins of many phages (including λ) mimic LexA, and their cleavage is likewise stimulated, inducing lytic development.

The repressor concentration is delicately poised. It is sufficient to control phage gene expression, but too much repressor impedes induction. Excess repressor synthesis is avoided because, although repressor stimulates transcription from pM at moderate concentrations, high repressor concentrations repress transcription from pM.

If a lysogenic cell is externally infected by another phage of the carried type, the repressor present in the cell binds to the operator sites on the entering phage and prevents it from initiating lytic development. The lysogenic cell is said to be immune. Repressor (and hence immunity) are specific. A cell lysogenic for λ can be lytically infected by unrelated phages such as Mu-1, and vice versa. Among the natural relatives of λ, there are a variety of repressor types; thus λ and 434, which insert at the same chromosomal site, make different repressors, and each phage makes plaques on lysogens of the other.

The circuitry has variations, even among close relatives of λ. Phage P22, for example, encodes, in addition to a repressor, an antirepressor protein that can bind to repressor and neutralize its activity. Antirepressor synthesis is turned off in lysogenic cells by a "maintenance" protein, Mnt, which represses transcription of the antirepressor gene. Phages such as P1 that lysogenize by plasmid formation also control most of their genes by repression. Their problem is inherently more complex than that of λ or Mu-1, because the lysogen must express not only the repressor gene but also those genes needed for plasmid replication and partitioning.

Although there are many examples of latency among eukaryotic viruses, there are few close analogues of lysogeny. Retroviruses, for example, insert their DNA and can replicate as part of a host chromosome, but they carry out a productive cycle of infection while inserted. When retroviruses are carried permanently in an inactive state, it is generally because either the virus has mutated to a defective form or the particular host cell type is unable to support a complete infectious cycle. The closest thing to lysogeny occurs with some of the nonintegrated viruses, such as papilloma and members of the Herpes family. In both these cases there seems to be a programmed potentiality either to be carried as a plasmid (analogous to P1) or to initiate a productive infection that culminates in virus liberation and cell death. In some of the herpesviruses this occurs in nondividing nerve cells, where the plasmid form need not replicate to persist; however, Epstein–Barr virus can establish a carried state in replicating lymphocytes.

Repressor Control During Lysogenization

Infection by a temperate phage has an indeterminate outcome. Some cells in the culture become lysogenic, whereas others are channeled into lytic development. If 100% of the cells gave a lytic response, the phage would be virulent. If 100% gave a lysogenic response, the phage would not form plaques and would not be readily recognizable as a phage. How these dual potentialities are achieved is best understood in λ. Chapter 8 covers present understanding of how the lysis/lysogeny decision is made in cells infected by λ. A major player is the cII protein. Mutations inactivating the *cII* gene strongly reduce the frequency of lysogenization, although the few lysogens formed are stable. *cII* is needed only for the establishment of lysogeny, not for its perpetuation.

cII is a transcriptional activator that coordinates the various aspects of lysogeny. To become lysogenic, a cell must not only establish repression of lytic genes, it must also integrate into the chromosome, and it must minimize the consequences of any lytic gene transcription initiated before the cell commits to lysogeny. There is no advantage to any one of these processes in absence of the others. It is futile to integrate into the chromosome if the cell will lyse for lack of repressor. Repression is established by transcription of the *cI* gene from a *cII*-dependent promoter within the *cII* gene. This promoter is activated only in the transient period during lysogenization. Once repression is established, the *cII* gene is itself repressed, and *cI* is transcribed only from the maintenance promoter pM.

Likewise, integrase is transcribed from a *cII*-dependent promoter (pI) with a start site within *xis*. Thus integrase, but not excisionase, is made in large amounts during lysogenization. Following induction, on the other hand, *int* and *xis* are coordinately expressed in the pL transcript.

A third *cII*-dependent promoter, pAQ, makes an antisense transcript that opposes expression of the *Q* gene, whose product stimulates late gene expression through antitermination. Therefore late gene products, which could kill and lyse the cell, are not made by cells expressing *cII*. Hence cells expressing *cII*, and only those cells, are effectively channeled into lysogeny. Thus in λ the indeterminacy of the outcome of infection depends completely on fluctuations in cII production. Other phages may employ very different mechanisms but, in order to be temperate, they must achieve the same end result.

REFERENCES

1. Groisman, E. A. and M. J. Casadaban. 1987. In vivo DNA cloning with mini-Mu replicon cosmid and a helper lambda phage. Gene 51:77–84.
2. Groisman, E. A., B. A. Castilho. and M. J. Casadaban. 1984. In vivo DNA cloning and adjacent gene fusing with a mini-Mu-lac bacteriophage containing a plasmid replicon. Proc. Natl. Acad. Sci. USA 81:1480–1483.
3. Hayashi,T., K. Makino, M. Ohnishi, K. Kurokawa, K. Ishii, K. Yokoyama, C.-G. Han, E. Ohtsubo, K. Nakayama,T. Murata, M. Tanaka,T. Tobe,T. Iida, H. Takami,T. Honda, C.Sasakawa, N. Ogasawara,T.Yansunaga, S. Kuhara,T. Shiiba, M. Hattow, and N. Shinagawa. 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* 0157:H7 and genomic comparisons with a laboratory strain of K-12. DNA Res. 8:11–22.
4. Jacob, F. and J. Monod. 1961. Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. 3:318–356.
5. Lwoff, A. and A. Gutmann. 1950. Recherches sur un *Bacillus megatherium* lysogene. Ann. Inst. Pasteur (Paris) 78:711–739.
6. Sternberg, N., D. Hamilton, S. Austin, M. Yarmolinsky. and R. Hoess. 1980. Site-specific recombination and its role in the life cycle of bacteriophage P1. Cold Spring Harbor Symp. Quant. Biol. 45:297-309.
7. Thyagarajan, B., E. C. Olivares,R. P. Hollis, O. S. Ginsburg. and M. P. Calos. 2001. Site-specific genomic integration in mammalian cells mediated by phage phi31 integrase. Mol. Cell. Biol. 21:3926–3934.
8. Young, S. M., Jr., D. M. McCarty, N. Degtyareva. and R. J. Samulski. 2000. Roles of adeno-associated virus Rep protein and human chromosome 19 in site-specific recombination. J.Virol. 74:3953–3966.

8

Gene Regulatory Circuitry of Phage λ

JOHN W. LITTLE

The gene regulatory circuitry of phage λ was the first complex regulatory circuit to be analyzed in molecular detail. λ was chosen as a model system for this purpose because it displays a wide range of interesting regulatory properties. It can adopt two alternative life-styles, the lytic and lysogenic states, and it can switch from the lysogenic state to the lytic state in the process of prophage induction, often called the "genetic switch" (13, 28). A choice is made between these two alternative pathways soon after infection; the lysogenic state, in particular, is highly stable, and the switch can be highly efficient. It was clear early on that these properties would be of general relevance to other organisms, including higher eukaryotes. λ proved a fortunate choice as an experimental system: its genetics is easy, and many of the regulatory proteins are also biochemically tractable, allowing a highly productive interplay between genetics and biochemistry. As a result of work by many investigators, the proteins and *cis*-acting sites involved in these regulatory events are well known and intensively studied, and most of their regulatory interactions are known in great mechanistic detail. Hence, λ remains the best-understood complex regulatory system. At the same time, new surprises continue to appear (12).

The primary purpose of this chapter is to outline the molecular events controlling the various regulatory decisions and states in the λ life cycle; to indicate areas of interest for future investigation; and to draw parallels between events in λ and other systems.

λ Background

After infecting its host, λ can follow either of two mutually exclusive pathways (Figure 8-1). In the *lytic* pathway, which resembles the life cycle of nearly all bacteriophages, the infected cell follows a temporal program of gene expression. Early gene products allow replication of viral DNA; late gene products include virion structural proteins, leading to assembly of mature virions, and proteins that cause cell lysis. In the *lysogenic* pathway, expression of the lytic genes is blocked by a protein called CI (also known as λ repressor); the viral DNA is integrated into the genome of the host by the action of the phage Int protein and the host IHF protein. In this form the phage is termed a prophage, and the viral DNA is replicated as part of the host DNA, offering the phage an alternative way to propagate. In the lysogenic state, CI continues to be expressed, and it represses expression of the lytic genes. This regulatory state is extremely stable, but it can be switched to the lytic state when the host SOS response is induced.

Hence, λ exhibits many interesting regulatory features, most of which were investigated for the first time in λ. Understanding these requires answers to several central questions. First, a regulatory decision is made between two alternative and mutually exclusive pathways. How is that decision made, and how does it respond to cellular physiology? Second, the lysogenic pathway leads to a highly stable regulatory state. How is that state stabilized? Third, the lysogenic state can break down. What are the events that lead to switching, and how do they fit with the needs of the phage? It is simplest to begin with the second question.

Stabilization of Regulatory States

As a starting point for understanding the regulatory events, we consider the regulatory states of the λ system in light of two rules involving two central regulatory proteins, the CI and Cro proteins: If the cells contain CI and no Cro, they are in the lysogenic state, and they continue to make CI and no Cro. If, on the other hand, the cells contain Cro but no CI, they are in the lytic state, and they continue to make Cro and no CI. The primary exception to this latter rule occurs during the lysis–lysogeny decision (see below). According to these rules, each regulatory state perpetuates itself, as we now describe. For a detailed review see Ptashne (28).

Most of the regulatory events that stabilize the regulatory states occur at a complex regulatory region termed the O_R region (Figure 8-2D). As will become clear below, a newly discovered long-range interaction occurs between O_R and another site, the O_L region (Figure 8-2B),

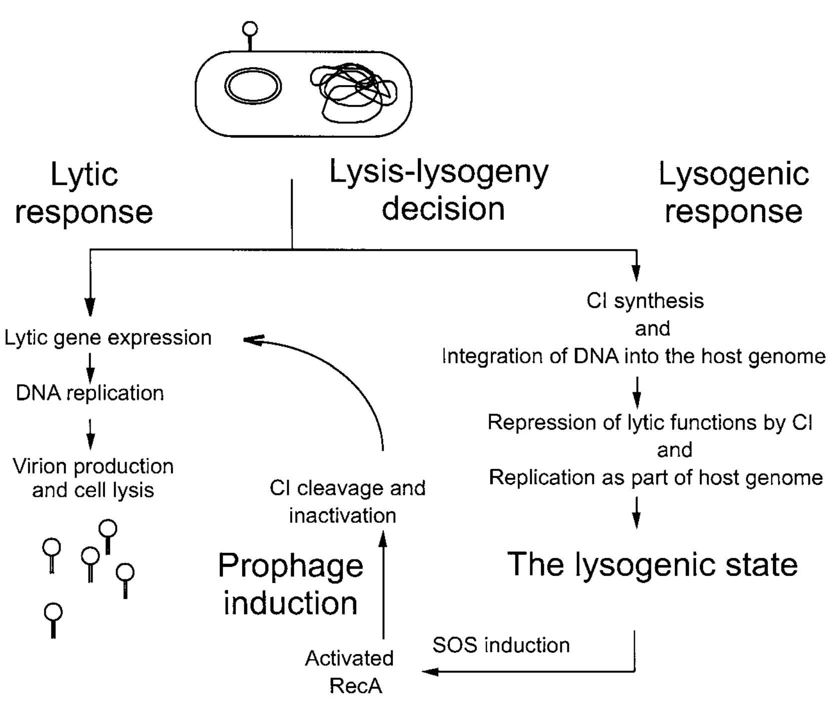

Lytic response
Lysis-lysogeny decision
Lysogenic response
Lytic gene expression
DNA replication
Virion production and cell lysis
CI synthesis and Integration of DNA into the host genome
Repression of lytic functions by CI and Replication as part of host genome
The lysogenic state
CI cleavage and inactivation
Prophage induction
Activated RecA
SOS induction

Figure 8-1 Phage λ life cycle. At the top is depicted an infected cell; when linear λ DNA is injected it cyclizes through cohesive termini, giving a circular molecule depicted at the left. The lysis-lysogeny decision is made 10–15 minutes after infection. See text for details.

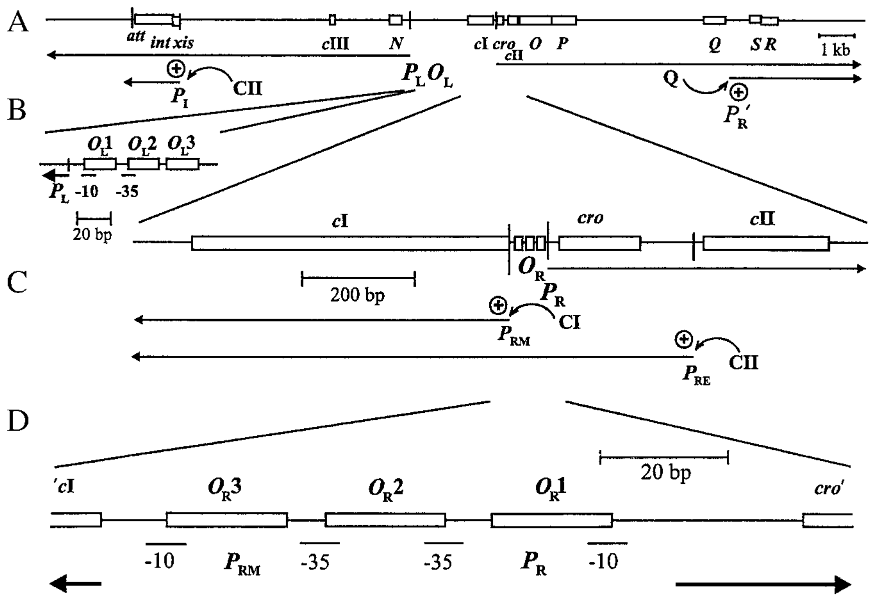

A
att int xis
cIII
N
cI cro cII O P
Q S R
1 kb
$P_L O_L$
P_I
CII
Q
P_R'
B
O_L1 O_L2 O_L3
P_L -10 -35
20 bp
C
cI
cro
cII
O_R
P_R
200 bp
P_{RM}
CI
P_{RE}
CII
D
'cI
O_R3
O_R2
O_R1
20 bp
cro'
-10
P_{RM}
-35
-35
P_R
-10

Figure 8-2 Maps of phage λ. The maps are to scale, as indicated; not all genes are shown. A: Map of the right half of the λ genome. Genes mentioned in the text are shown; in addition, *att* is the site of site-specific recombination for integrating and excising the prophage; *xis* is required for prophage excision; S and R are involved in cell lysis; and Q is an antitermination protein for transcripts initiating at the late lytic promoter P_R', which otherwise terminate before *S*. Transcripts from the P_I, P_L, P_R, and P_R' (also known as pI, pL, pR, and pR′, respectively) promoters are shown. B: Close-up of the O_L region; this region is organized similarly to the O_R region (also known as oL and oR, respectively), except that it lacks a oppositely directed promoter analogous to P_{RM}. C: Close-up of the region around O_R. Locations of the P_R, P_{RM}, and P_{RE} promoters and their transcripts are shown. *cro* is translated from the P_R transcript, not from the P_{RE} transcript. D: Extreme close-up of the O_R region. Only the initial portions of the *cI* and *cro* genes are shown. The P_{RM} message starts with the AUG start codon of *cI*, and lacks a Shine-Dalgarno sequence. See figure 27-2 for a more detailed map of the phage λ genome.

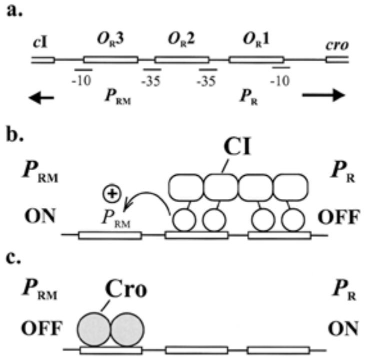


Figure 8-3 Molecular basis for bistability of λ circuitry. A: The O_R region is repeated from Figure 8-2D. B: Occupancy pattern of O_R operators at moderate CI levels. CI binds as a dimer, and dimers at O_R1 and O_R2 make cooperative interactions. Contrary to this cartoon, structural evidence suggests that both subunits of a dimer touch the other dimer; evidence with LexA suggests that the "hinge" region connecting the dimerization and DNA-binding domains is shorter than implied here. C: Occupancy patterns of O_R operators at moderate levels of Cro.

and this interaction somewhat complicates our understanding of the events at O_R. Nonetheless, the events at the O_R region determine the regulatory state of the system.

The O_R region contains five sites to which proteins bind: two promoters, termed P_{RM} and P_R, and three binding sites for regulatory proteins, O_R3, O_R2 and O_R1. The two regulatory proteins, CI and Cro, bind to these three sites, but the consequences of binding differ, as detailed below.

P_{RM} and P_R are expressed during the lysogenic state and the lytic state, respectively. Each promoter expresses one of the two regulatory proteins. Expression of P_{RM} leads to synthesis of CI. Expression of P_R leads to synthesis of Cro, as well as to synthesis of several lytic gene products (O, P and Q), which we will not consider further. Hence, the rules given above can be restated: CI allows expression of P_{RM} but not P_R; Cro allows expression of P_R but not P_{RM}. That is, in each case the regulatory protein expressed by the promoter serves to stabilize the regulatory state that allows its expression.

To understand the reasons for this, we next consider the detailed organization of the O_R region (Figure 8-2D and Figure 8-3). As stated, both Cro and CI bind to three binding sites in this interval. However, binding of the two proteins differs in several ways, and has markedly differing consequences. Cro protein, which is simpler than CI, can bind to O_R3, O_R2, and O_R1, but it binds most tightly to O_R3. Hence, at moderate Cro levels, O_R3 is occupied but O_R2 and O_R1 are free (Figure 8-3C). When Cro binds to O_R3, it completely represses P_{RM}, since this promoter overlaps O_R3. However, in this configuration Cro has no effect on P_R, which does not overlap O_R3. As Cro levels rise further, Cro can bind to O_R2 and/or O_R1, repressing P_R; hence the level of Cro is self-limiting due to negative autoregulation.

The case of CI is more complicated. CI binds tightly to O_R1 but weakly to O_R2 and O_R3. CI has an additional feature, however, which leads to occupancy of O_R2: CI binds cooperatively to adjacent binding sites. That is, occupancy of one site increases the affinity of the protein for an adjacent site roughly 200-fold (5). Since O_R1 is a tight binding site, this leads to much tighter binding to O_R2. In addition, cooperative binding by λ CI is "alternate pairwise": some feature of the protein–DNA complex prevents a CI dimer at a site from interacting simultaneously with dimers at two flanking sites. When O_R1 is mutated, CI can bind cooperatively to O_R2 and O_R3, and the favorable free energy for this interaction is about the same as for cooperative binding to O_R1 and O_R2. However, a dimer at O_R2 cannot interact both with ones at O_R1 and O_R3; since binding to O_R1 is intrinsically tighter than that to O_R3, occupancy of O_R1 and O_R2 is favored (Figure 8-3B).

Occupancy of O_R1 and O_R2 has two different consequences. First, it leads to repression of P_R, which overlaps these two sites. Second, it leads to *activation* of P_{RM}: when O_R2 is occupied by CI, P_{RM} is about 10 times stronger than when O_R2 is vacant. This *positive autoregulation* of CI expression strongly reinforces the lysogenic state—that is, if CI is present it continues to be made, and it stimulates its own expression. If CI levels rise above a certain level, CI begins to bind to O_R3, partially shutting off P_{RM}. It was long believed that this negative autoregulation was unimportant in a λ lysogen, but recent evidence shows that such negative feedback does occur in λ lysogens, and that it involves a long-range interaction with the O_L region (see below).

These observations suffice to explain the rules given above: CI allows expression of P_{RM} by stimulating it, and prevents expression of P_R by steric hindrance. At moderate levels, Cro allows expression of P_R because it does not bind to sites overlapping P_R, but represses P_{RM} by binding to a site overlapping this promoter (Figure 8-3).

Although this account describes how the lysogenic state is stabilized, it does not explain the way in which it is established. Immediately after infection of a host cell by λ, neither CI nor Cro is present. Hence P_R will be expressed

(P_{RM} is a weak promoter in the absence of CI), leading to Cro expression. How can a cell ever achieve the regulatory state of CI present and Cro absent? The answer, described further below, is that CI can be expressed from a second promoter, termed P_{RE} (Figure 8-2C), which is controlled in a different way in response to the physiology of the cell. If an infected cell is destined to follow the lysogenic pathway, P_{RE} is expressed, leading to very high levels of CI; this overrides the controls described above and shuts off expression of both P_R and P_{RM} for several generations, until CI levels drop due to dilution of cell contents during growth, allowing expression of P_{RM}.

Unlike the lysogenic state, the lytic state cannot be and does not need to be highly stable, since cells following the lytic pathway lyse in about an hour. In certain mutants that block lytic growth, however, the lytic state can be stable (10, 26). In most studies, these mutants had a temperature-sensitive *c*I allele, *c*I857, and defects in the N antitermination protein and in a DNA-replication protein (O or P). When lysogens are grown to 40°C to inactivate CI, then shifted to 30°C, they remain in an "anti-immune" state for many generations; this state is not observed if the prophage also has a *cro* mutation. Hence, in this state, Cro has established dominance over CI. The anti-immune state is not completely stable; after many generations, a fraction of the cells switch back to the immune state.

Prophage Induction: The Switch from the Lysogenic State to the Lytic State

Although the lysogenic state is extremely stable, lysogens can switch to the lytic state. They do so by taking advantage of a host regulatory system termed the SOS response (11, 21). In response to conditions that damage DNA or inhibit DNA replication, perhaps 40 SOS genes are expressed at elevated levels. The products of these genes help the cell cope with the DNA damage in various ways. The SOS response involves the interplay of two regulatory proteins: RecA and LexA. During normal growth, LexA partially represses the SOS genes. After inducing treatments, the RecA protein is activated to a form, termed RecA*, which catalyzes the specific proteolytic cleavage of LexA. This reaction is unusual, in that RecA acts indirectly as a coprotease to stimulate an inherent self-cleavage activity of LexA (19, 20). Recent crystal structures of several mutant LexA proteins indicate that this activity of LexA is regulated by a conformational change (23). In the noncleavable conformation, the cleavage site is distant from the active site that catalyzes cleavage; it is likely that LexA is usually in this conformation. In the cleavable conformation, by contrast, the cleavage site and active site are juxtaposed. It is likely that RecA somehow stabilizes this conformation.

λ CI is cleaved in an entirely parallel reaction. Structural data (2) show that the active site is closely similar to that of LexA; the cleavage site is not present in these structures. CI has a slow self-cleavage activity, and RecA stimulates this reaction. Hence, after conditions that induce the SOS response, λ CI is also cleaved. The SOS response is a rapid response; cleavage of LexA is nearly complete by 5 minutes after UV irradiation. Since cleavage of CI is far slower, it takes about 30 minutes for CI levels to drop low enough to derepress the lytic P_R promoter; once P_R is derepressed, Cro is made, which then binds to O_R3, probably preventing further synthesis of CI and making the switch irreversible.

Cleavage of CI is much slower than that of LexA, both because dimers of CI are resistant to cleavage and because the rate is intrinsically slower. It is believed that this disparity in rates reflects the biological niches of the two repressors: LexA is designed for a rapid and reversible response to inducing treatments, while CI has likely evolved to allow efficient prophage induction only at doses of DNA damage that are likely to kill the host cell (20).

In addition to prophage induction caused by overt inducing treatments, cultures of lysogens contain some free phage particles. These are released in a process that also requires the SOS response, as shown by two lines of evidence: very low levels of free phage are seen in cultures of host *recA* mutants, and in *recA*$^+$ lysogens of λ containing a *c*I *ind*$^-$ mutation, which blocks CI cleavage (22). This process is often termed spontaneous induction. Although it can be argued that this is a misnomer since the investigator does nothing to induce the process, the process does result from induction of the SOS response, and the term reflects the belief that the switch to the lytic state results from induction of the SOS pathway, whatever the stimulus may be.

Recent evidence points toward a likely source for an inducing signal (7). In normally growing cells, replication forks frequently encounter single-strand breaks or noncoding lesions, leading to breakdown of the replisome. It is plausible that such events generate a transient activation of RecA; in this view, rare cells have enough damage to keep RecA activated long enough to cleave CI, leading to switching.

Role of O_L

Recent evidence indicates that the O_R region is not a self-contained regulatory module. It has long been known that CI and Cro both bind to a second regulatory region, termed the O_L region. In the lysogenic state, binding of CI to O_L is necessary to repress expression of the lytic P_L promoter, which codes for the N protein, an antitermination protein essential for the lytic program of expression. Several new lines of evidence indicate a second role for binding of CI to O_L: CI bound to O_L can cooperatively interact with CI bound at O_R (Figure 8-4).

First, in P_R::*lacZ* fusions, presence of O_L at a site distal to *lacZ* leads to increased repression of *lacZ* (31). This interaction requires only two operator sites at O_R and two sites

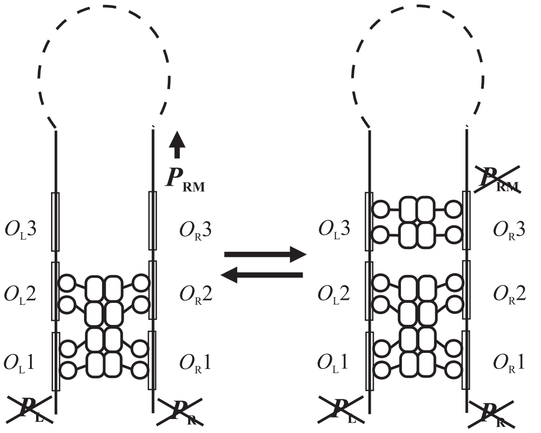


Figure 8-4 O_R:O_L interaction. One possible set of interactions between CI bound at O_L and O_R. On the left is shown an interaction involving dimers bound at each of four sites. The intervening 2.4 kb form a loop (indicated by the dashed arc). This structure would form from a non-looped structure (not shown) by cooperative interactions between tetramers at each site. Both P_L and P_R are presumed to be repressed in this structure, as they would be in the absence of looping when CI binds. On the right is a complex containing additional dimers bound to O_R3 and O_L3; in this complex, P_{RM} is repressed. Other complexes may also form; for instance, the O_L region might be inverted from the orientation.

at O_L. Second, CI can form loops between O_L and O_R, as observed by electron microscopy (31). Third, the lysogenic P_{RM} promoter can also be repressed more efficiently by CI when O_L is distal to the *lac*Z reporter gene; this effect is not seen in the presence of an O_R3 mutation blocking CI binding, indicating that P_{RM} repression involves CI binding to O_R3 (8). In addition, lysogens of this O_R3 mutant are more difficult to induce by UV light than wild-type. It is suggested that a complex forms involving dimer–dimer interaction between CI at O_R3 and at O_L3. Fourth, O_L3 mutations make it more difficult to induce λ by UV, indicating a role for O_L3 and further supporting the model of an O_R3:O_L3 interaction (C. B. Michalowski and J. W. L, unpublished data).

The primary interaction between O_R and O_L is likely mediated by protein–protein interactions between tetramers bound to the two regions, at O_R1 and O_R2 and at O_L1 and O_L2, respectively (Figure 8-4). It is known that CI can form octamers in solution, and a recent crystal structure (3) of the C-terminal domain shows the structure of the octamer. In this view, tetramer–tetramer interaction holds O_R and O_L together; weak binding of dimers to O_R3 and O_L3 is strengthened by cooperative interaction between the two dimers (Figure 8-4). However, it remains plausible that several different types of complexes can form; for instance, perhaps O_R can approach O_L in two orientations, which would likely allow different complexes to form. In addition, if multiple complexes exist, they may interconvert readily.

Whatever the details of this interaction, it has several regulatory consequences, which are beginning to be understood. First, since it is a cooperative interaction, it probably strengthens binding of CI to O_R and to O_L, conferring more complete repression of the lytic P_R and P_L promoters during the lysogenic state. Second, as stated it leads to substantial negative autoregulation of *c*I at the levels of CI found in a lysogen, ensuring a more constant level of CI. Finally, it is tempting to speculate that cooperative binding provides a way to coordinate derepression of the lytic promoters during the process of prophage induction. In this view, as CI levels drop due to cleavage, when the long-range interaction was broken, binding of CI to O_R and O_L would also be weakened, causing CI to fall off both sites at the same time, ensuring simultaneous derepression of P_R and P_L. One line of evidence compatible with this is the finding that the phage $\lambda O_R3'23'$ has a substantially reduced burst size after UV induction (22), despite normal lytic growth; perhaps the alterations of both O_R1 and O_R3 in this variant disrupt the normal interaction between O_R and O_L.

Lysis–Lysogeny Decision

As described above, CI must be present to stimulate expression of P_{RM} in the lysogenic state. In a cell that is following the lysogenic pathway, the initial burst of CI is provided by expression of another promoter, P_{RE} (9, 16). Regulation of this promoter involves physiological events that are not fully understood. The direct positive regulator of P_{RE} is the phage protein CII, which binds at P_{RE} (Figure 8-2C) and activates its expression. The level of CII in the cell is in turn under complex control. CII is degraded by a host protease, FtsH (or HflB) (18, 32), and the action of FtsH is antagonized by another phage protein termed CIII. CIII apparently acts as a competitive inhibitor of FtsH (14). It is unclear whether physiological conditions also influence the activity of FtsH. Infection of cells with multiple phages allows more efficient lysogenization, and it is believed that the high gene dosage provides more CII and CIII, stabilizing CII. Even under these conditions, however, not all the infected cells follow the lysogenic pathway.

The lysis–lysogeny decision is made about 10–15 minutes after infection (9). During this time, expression of the early lytic genes *N*, *O*, and *P*, and the *c*II and *c*III genes, has occurred. It is likely that the decision is only made after this time to allow CII, CIII, FtsH, and perhaps another factor, HflKC (see below), to interact in such a way as to allow sensing of the physiological state of the cell. In a cell following the lysogenic pathway, CII also activates expression of the Int protein, which is required for integration of the viral genome into that of the host.

The lysis–lysogeny decision has been simulated by a computer model (25) that involves all the known components (but not the O_L:O_R interaction). This simulation mimics the in vivo behavior of multiply infected cells, in that not all simulations lead to the lysogenic response. The computer model is a "stochastic" model, in which expression of promoters, translation and degradation of mRNAs, and degradation of proteins occurs not at a fixed rate but with a given probability per unit time. Different simulations give widely varying levels of gene products; in particular, the level of CII varies markedly from one run to another. The authors ascribe the variable response of the model to this stochastic behavior, and note that this type of stochastic behavior should have the most pronounced effects in cells with small numbers of molecules.

Biochemical analysis of CII degradation has been complicated by several factors. First, *ftsH* is an essential gene, complicating genetic analysis. Second, FtsH is a membrane-bound protein, making biochemical analysis difficult. Third, its activity in vivo is apparently modulated by interaction with the HflKC complex. Mutations inactivating the HflKC complex also favor lysogeny, suggesting that its interaction with FtsH modulates the activity of FtsH. Fourth, HflKC is also membrane-bound, and the portion not embedded in the membrane lies primarily in the periplasmic space. This greatly complicates in vitro reconstruction of its interaction with FtsH, since the nonmembrane portions of HflKC and FtsH lie on opposite sides of the inner membrane.

What Is Essential, and What Is Refinement?

Are the detailed features of the λ regulatory circuitry essential for its proper operation? Or are they refinements to a basic ground plan? This issue is important for evolution of complex gene regulatory circuitry. If a circuit needs all its "bells and whistles" to work at all, then one would expect that it would arise during evolution only with difficulty. If, by contrast, a stripped-down version of the regulatory circuit would suffice in the initial stages of evolution, and the refinements were added later, the circuit would seem far more likely to arise (22, 29). In this two-step model, the initial version would have to offer a selective advantage, such as the lysogenic life-style; the later refinements would confer optimal behavior.

Relatively limited analysis of other lambdoid phages, such as 434 and HK022, suggests that these phages have similar behaviors, such as positive autoregulation of CI, cooperative DNA binding by CI, and differential occupancies of the O_R operators (6, 28, 34). With this limited data set, we can infer that these properties are generally advantageous to a lambdoid phage, and have been selected during evolution. In addition, since the repressors have different specificities, at least some of these features likely evolved independently in different immunity specificities.

Several recent lines of evidence suggest that a two-step model is compatible with the properties of the λ circuitry. First, the differential affinities of Cro and CI for the various operators in O_R are not an *essential* part of λ regulation, since O_R1 and O_R3 can be made identical and the phage exhibits near-normal behavior (22). These two sites differ in three positions. Three variant phages were made in which the operators at the sites of both O_R1 and O_R3 have the same sequence, either those of O_R1, O_R3, or a hybrid site differing from O_R3 by one change. All these variants are able to grow lytically, readily form stable lysogens, and lysogens are readily induced by UV. They show quantitative differences, but their qualitative operation is like that of the wild-type. At the mechanistic level, the behavior of these phages can be rationalized in terms of the known properties of CI, Cro, and the O_R region. In particular, in a lysogen it is likely that CI frequently alternates between cooperative binding to O_R1 and O_R2 (as in the wild-type, Figure 8-3B) and to O_R2 and O_R3 (not shown), repressing both promoters; the sum of these two occupancy patterns is P_{RM} ON and P_R OFF. In any case, although the differential affinities shown by CI and Cro for the O_R operators appear to make the circuit work better, they are likely a refinement.

This is an example of a property termed "robustness," which means that variations in the parameters of a system still allow its normal or near-normal function. Robustness facilitates a two-step pathway for evolution, since it permits a wider range of circuits to arise initially (1, 22).

Second, cooperative binding of CI is not essential. We have found that a cooperativity mutant in *c*I (Y210N), in combination with two mutations in the O_R region, exhibits near-normal behavior (A. C. Watson and J. W. L., unpublished data). The Y210N mutant has been shown to be defective in cooperative binding to adjacent sites in vivo (35), and presumably it blocks the interaction between O_L and O_R as well. Again, these data strongly suggest that cooperativity is a refinement rather than an essential feature.

Third, positive autoregulation by CI at P_{RM} is not essential (C. B. Michalowski and J. W. L., unpublished data). Two mutations in *c*I, D38R and D38N, were previously shown to prevent P_{RM} activation in a model system. When crossed onto a phage, D38R does not allow lysogenization, but it can be suppressed by mutations in P_{RM}. A phage carrying D38N can form extremely unstable lysogens, but changes in P_{RM} allow it to form lysogens about as stable as wild-type. These P_{RM} mutations do not restore positive control by D38R or D38N mutant CI. This analysis suggests that positive autoregulation of *c*I expression is also a refinement.

With respect to the role of CII in regulation, at least one phage, 21, does not have an identifiable P_I promoter for CII-dependent *int* expression (36). In addition, in λ the CIII function is dispensable in the presence of a particular CII mutant allele, *can-1* (16). These findings suggest that aspects of the CII system are refinements as well.

Generalizations to Other Regulatory Systems

Many of the regulatory principles described above were first established by studies on λ, and all are generally applicable to a wide range of regulatory circuits.

Establishment and Maintenance

The principle that a regulatory decision is first made in an establishment phase, then ratified and made permanent in a maintenance phase, is observed in a variety of systems. Perhaps the best studied case is that of *Sxl* in *Drosophila*. In the establishment phase, *Sxl* is expressed from the *Pe* promoter in females but not in males (17). Translation of the *Pe* transcript makes active Sxl protein. In the maintenance phase, the *Sxl* gene is expressed from a different promoter, the *Pm* promoter, both in males and in females; however, the following positive autoregulatory loop ensures that Sxl protein is made only in females. The Sxl protein is required for proper splicing of the latter mRNA, and only in its presence can the *Pm* transcript be spliced properly. Accordingly, in females the presence of Sxl protein made in the establishment phase allows more Sxl protein to be made in the maintenance phase; in males, the *Pm* transcript cannot be properly spliced, and no Sxl protein is made. As another example, *Drosophila* proteins in the Polycomb group act to maintain patterns of gene expression established by other mechanisms, doing so by maintaining repressive chromatin structures (4).

Multiple Promoters

The case of P_{RE} and P_{RM} was the first instance in which two different promoters were shown to drive expression of the same gene (30). Importantly, expression of each of these promoters has a different regulatory meaning, as described above. Perhaps the best-analyzed eukaryotic system that reflects these principles is that controlling expression of "pair-rule" genes in *Drosophila*. In a well-studied case, the *eve* gene is expressed in seven stripes in the early embryo (33). Each of these stripes appears to be under separate control, and several of them are regulated by a discrete region of the *eve* regulatory region; for instance, the stripe 2 enhancer contains binding sites for four proteins, most of which are products of "gap" genes, and the proper combination of these proteins is found only in the region in which stripe 2 is expressed (33). This stripe 2 enhancer can be separated from the rest of the *eve* upstream region, and it functions autonomously.

In this system, then, there are seven different sets of inputs that give the common output of *eve* expression. These seven inputs correspond to conditions occurring in seven regions or zones along the anterior–posterior axis of the embryo. The common output establishes seven zones of expression of the same regulatory protein. Concurrent regulation of several other pair-rule genes establishes a pattern that repeats seven times in the embryo, dividing it up into seven regions that then develop further in a pattern that is in many ways equivalent in each region.

The logic of the *eve* system is the same OR logic as that first established in λ: "Express the gene if condition 1 OR condition 2 OR . . . condition 7 is met." That is, any of seven different conditions suffices to turn the gene on. This logic can be contrasted with another well-analyzed example, that of *HO* regulation in *Saccharomyces cerevisiae*, in which the logic is "express the gene only if condition 1 AND condition 2 AND condition 3 are met" (15).

Alternative Stable Regulatory States

Stable regulatory states are relatively rare in prokaryotes; most regulatory systems are rapidly reversible, allowing dynamic responses to changes in the environment. In Metazoa, stable regulatory states are universally used to establish and maintain particular cell types. The mechanistic details in higher eukaryotes are certainly more complex, in part since they involve heritable states of chromatin.

Threshold Behavior

Prophage induction in phage λ exhibits threshold behavior; below a certain threshold level of DNA damage this process is rather inefficient, while above the threshold most or all of the cells switch their state (e.g., 22). Many examples of threshold behavior are known in eukaryotic regulation. Once again, the best-analyzed examples are in *Drosophila*. During early embryogenesis, several gap genes are expressed only above a certain threshold of maternal gene products (24, 27); for instance, zygotic Hb is made only above a certain threshold concentration of Bcd protein.

Conclusions

The λ regulatory circuitry is well understood at the mechanistic level. The behavior of mutants is generally consistent with the in vitro behavior of the regulatory proteins and their targets, and with mutant variants of these elements. As noted above, a relatively successful computer simulation has been developed of the lysis–lysogeny decision that relies on quantitative estimates for about 40 different parameters. This simulation can reproduce the lysogenization frequency of cells infected with various multiplicities of λ. On the other hand, it does not predict the large overshoot of CI synthesis from P_{RE}, indicating the need for further refinement of the model. Nonetheless, the success of this approach suggests that we know most of the elements involved in the lysis–lysogeny decision, and can make reasonable quantitative inferences about their properties and interactions.

Several fruitful avenues remain for future research. The CI-mediated interaction between O_R and O_L needs far better characterization, as regards both the nature of the complexes formed and the functional consequences of this interaction. Further analysis of events in the lysis–lysogeny decision is important; among these are the detailed mechanisms of CII cleavage by FtsH, the modulation of FtsH activity by other cellular proteins such as HflKC, the inhibitory action of CIII, and particularly the relationship between cellular physiology and the efficiency of CII cleavage. Finally, further analysis of simplified versions of the λ circuitry, and comparison with the properties of other lambdoid phage circuits, should help lend insights into mechanisms by which complex circuits could plausibly evolve.

References

1. Barkai, N. and S. Leibler. 1997. Robustness in simple biochemical networks. Nature 387:913–917.
2. Bell, C. E., P. Frescura, A. Hochschild, and M. Lewis, 2000. Crystal structure of the lambda repressor C-terminal domain provides a model for cooperative operator binding. Cell 101:801–811.
3. Bell, C. E. and M. Lewis. 2001. Crystal structure of the lambda repressor C-terminal domain octamer. J. Mol. Biol. 314:1127–1136.
4. Brock, H. W. and M. Van Lohuizen. 2001. The Polycomb group—no longer an exclusive club? Curr. Opin. Genet. Dev. 11:175–181.
5. Burz, D. S. and G. K. Ackers. 1994. Single-site mutations in the C-terminal domain of bacteriophage lambda *c*I repressor alter cooperative interactions between dimers adjacently bound to O_R. Biochemistry 33:8406–8416.
6. Bushman, F. D. 1993. The bacteriophage 434 right operator. Roles of O_R1, O_R2 and O_R3. J. Mol. Biol. 230:28–40.
7. Cox, M. M., M. F. Goodman, K. N. Kreuzer, D. J. Sherratt, S. J. Sandler, and K. J. Marians. 2000. The importance of repairing stalled replication forks. Nature 404:37–41.
8. Dodd, I. B., A. J. Perkins, D. Tsemitsidis, and J. B. Egan. 2001. Octamerization of lambda CI repressor is needed for effective repression of P(RM) and efficient switching from lysogeny. Genes Dev. 15:3013–3022.
9. Echols, H. 1986. Bacteriophage λ development: temporal switches and the choice of lysis or lysogeny. Annu. Rev. Genet. 2:26–30.
10. Eisen, H., P. Brachet, L. Pereira da Silva, and F. Jacob. 1970. Regulation of repressor expression in λ. Proc. Natl. Acad. Sci. USA 66:855–862.
11. Friedberg, E. C., G. C. Walker, and W. Siede. 1995. DNA Repair and Mutagenesis. American Society of Microbiology Press, Washington, D.C.
12. Friedman, D. I. and D. L. Court. 2001. Bacteriophage lambda: alive and well and still doing its thing. Curr. Opin. Microbiol. 4:201–207.
13. Hendrix, R. W., J. W. Roberts, F. W. Stahl, and R. A. Weisberg. 1983. Lambda II. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
14. Herman, C., D. Thévenet, R. D'Ari, and P. Bouloc. 1997. The HflB protease of *Escherichia coli* degrades its inhibitor λcIII. J. Bacteriol. 179:358–363.
15. Herskowitz, I. 1989. A regulatory hierarchy for cell specialization in yeast. Nature 342:749–757.
16. Herskowitz, I. and D. Hagen, D. 1980. The lysis–lysogeny decision of phage λ: Explicit programming and responsiveness. Annu. Rev. Genet. 14:399–445.
17. Keyes, L. N., T. W. Cline, and P. Schedl. 1992. The primary sex determination signal of *Drosophila* acts at the level of transcription. Cell 68:933–943.
18. Kihara, A., Y. Akiyama, and K. Ito. 1997. Host regulation of lysogenic decision in bacteriophage lambda: transmembrane modulation of FtsH (HflB), the cII degrading protease, by HflKC (HflA). Proc. Natl. Acad. Sci. USA 94:5544–5549.
19. Little, J. W. 1984. Autodigestion of lexA and phage lambda repressors. Proc. Natl. Acad. Sci. USA 81:1375–1379.
20. Little, J. W. 1993. LexA cleavage and other self-processing reactions. J. Bacteriol. 175:4943–4950.
21. Little, J. W. and D. W. Mount. 1982. The SOS regulatory system of *Escherichia coli*. Cell 29:11–22.
22. Little, J. W., D. P. Shepley, and D. W. Wert. 1999. Robustness of a gene regulatory circuit. EMBO J. 18:4299–4307.
23. Luo, Y., R. A. Pfuetzner, S. Mosimann, M. Paetzel, E. A. Frey, M. Cherney, B. Kim, J. W. Little, and N. C. J. Strynadka. 2001. Crystal structure of LexA: a conformational switch for regulation of self-cleavage. Cell 106:585–594.
24. Mannervik, M., Y. Nibu, H. Zhang, and M. Levine. 1999. Transcriptional coregulators in development. Science 284:606–609.
25. McAdams, H. H. and A. Arkin. 1997. Stochastic mechanisms in gene expression. Proc. Natl. Acad. Sci. USA 94:814–819.
26. Neubauer, Z. and E. Calef, E. 1970. Immunity phase-shift in defective lysogens: nonmutational hereditary change of early regulation of lambda prophage. J. Mol. Biol. 51:1–13.
27. Pankratz, M. J. and H. Jäckle. 1990. Making stripes in the *Drosophila* embryo. Trends Genet. 6:287–292.
28. Ptashne, M. 2004. A Genetic Switch: Phage Lamda Revisited. Cold spring Harbor Laboratory Press, Cold Spring Harbor, NY.
29. Ptashne, M. and A. Gann. 1998. Imposing specificity by localization: mechanism and evolvability. Curr. Biol. 8:R812–R822.
30. Reichardt, L. and A. D. Kaiser. 1971. Control of λ repressor synthesis. Proc. Natl. Acad. Sci. USA 68:2185–2189.
31. Révet, B., B. Von Wilcken-Bergmann, H. Bessert, A. Barker, and B. Müller-Hill. 1999. Four dimers of lambda repressor bound to two suitably spaced pairs of lambda operators form octamers and DNA loops over large distances. Curr. Biol. 9:151–154.

32. Shotland, Y., S. Koby, D. Teff, N. Mansur, D. A. Oren, K. Tatematsu, T. Tomoyasu, M. Kessel, B. Bukau, T. Ogura, and A. B. Oppenheim. 1997. Proteolysis of the phage λ CII regulatory protein by FtsH (HflB) of *Escherichia coli*. Mol. Microbiol. 24:1303–1310.
33. Small, S., A. Blair, and M. Levine. 1992. Regulation of *even-skipped* stripe 2 in the *Drosophila* embryo. EMBO J. 11:4047–4057.
34. Weisberg, R. A., M. E. Gottesman, R. M. Hendrix, and J. W. Little. 1999. Family values in the age of genomics: comparative analyses of temperate bacteriophage HK022. Annu. Rev. Genet. 33:565–602.
35. Whipple, F. W., N. H. Kuldell, L. A. Cheatham, and A. Hochschild. 1994. Specificity determinants for the interaction of λ repressor and P22 repressor dimers. Genes Dev. 8:1212–1223.
36. Wulff, D. L., Y. S. Ho, S. Powers, and M. Rosenberg. 1993. The *int* genes of bacteriophages P22 and lambda are regulated by different mechanisms. Mol. Microbiol. 9:261–271.

9

Regulation of λ Gene Expression by Transcription Termination and Antitermination

DAVID I. FRIEDMAN
DONALD L. COURT

Regulatory effects on gene expression at the level of transcription elongation are now recognized to be widely operative in both prokaryotes and eukaryotes (23, 97, 117, 123). Recognition of this mode of regulation can be dated to the 1969 *Nature* paper by Jeff Roberts (138). Not only did this work identify the *Escherichia coli* Rho termination protein and propose that there are specific transcription termination sites in the phage λ genome, but it led to the proposal that the λ N transcription regulation protein promotes gene expression by allowing RNA polymerase (RNA Pol) to transcend transcription terminators. The ideas first presented in this seminal paper have served as the basis for over 30 years of intensive work not only with phages but also with organisms on all levels of the evolutionary ladder. However, studies on λ and related lambdoid phages have continued to be prime sources of information about the molecular events influencing transcription elongation. It is the purpose of this chapter to update a previous review of this subject (48). We will present a short summary of what we consider essential material from that review as background and then a more extensive review that focuses primarily on subsequent work that we think has significantly extended understanding of the process of gene regulation by transcription termination and antitermination.

This chapter will focus on the lambdoid family, because the terminators and antiterminators of these phages have served as major tools in the study of transcription termination and antitermination. The impact of the work on these phages is clearly evident from the large number of review articles published on the subject since the last review in this series. We point to these reviews because they undoubtedly have different emphases than will be found in this chapter (29, 30, 50, 51, 67, 70, 74, 133, 137, 141, 177).

Note that we will use uppercase letters to indicate RNA (e.g., NUT) and lowercase italics (e.g., *nut*) to indicate DNA.

Prelude: The λ Story, What Came Before

Note that in this section we rely on the reviews listed above for many of the references. In some cases, where we feel it is warranted, further appropriate references will be included.

Transcription Patterns

This discussion is based on the information contained in figure 9-1. Similar transcription patterns have been observed with all lambdoid phages that have been examined. However, the small differences between λ and other members of the lambdoid family have provided important insights into the mechanism of regulatory processes that act during transcription elongation. These processes, transcription termination and antitermination, play important roles in the temporal regulation of transcription of the λ genome (figure 9-1). Transcription terminators, sequences that direct dissociation of the ternary RNA Pol–DNA–mRNA transcription complex are found at strategic positions on the genomes of lambdoid phages (indicated by t in figure 9-1). As in *E. coli*, these terminators can be either Rho-dependent or intrinsic (factor-independent). The N protein of λ, with *E. coli* Nus factors, modifies the transcribing RNA Pol at NUT sites, creating a termination-resistant transcription complex. The NUT sites, recognized on the RNA transcripts, are located approximately 35 and approximately 225 bases downstream, respectively, of the transcription starts sites of the early λ p_L and p_R promoters.

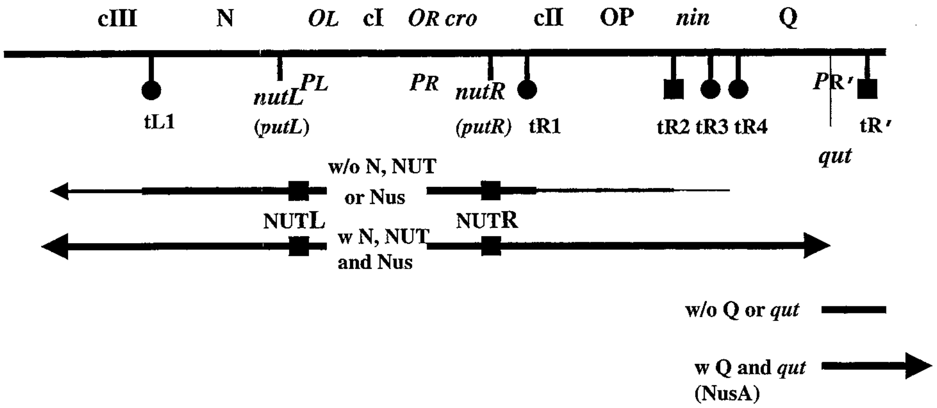


Figure 9-1 Genetic map of regulatory region of λ. (Map is not drawn to scale.) Shown are relevant promoters (*P*), operators (t), *nut* (*put*), and *qut*. Below are shown transcription patterns in the absence and presence of N and Nus factors and in the presence and absence of Q. Narrowing of arrows provides an indication of levels of partial transcription.

This N-directed antitermination allows full expression of genes downstream of transcription terminators in the p_L and p_R operons. The *Q* gene, encoding another antitermination factor, is 5.8 kb beyond the p_R promoter. Because there are at least four transcription terminators between the p_R promoter and the *Q* gene, N antitermination is required for Q expression. The Q antitermination protein acting at the late $p_{R'}$ promoter, in turn, is required for full expression of the phage late genes. As we will see, the requirements and mechanics of N and Q modification of the polymerase complex are quite different.

The N Antitermination Protein

The *N* gene (15) located immediately downstream of the early p_L promoter encodes a 12.2 kDa protein (46, 72) that, as discussed, acts at NUT sites to modify RNA Pol to a termination-resistant form. To a first approximation N proteins have specificity for their cognate NUT sequence, although when overexpressed in vivo, N proteins can act at noncognate NUT sites (47, 150).

Genetic studies suggested an interaction between N and RNA Pol. *E. coli* carrying a mutation (*ron*) in the *rpoB* gene (encoding the β subunit of RNA Pol) fails to support growth of λ with certain mutations, *mar*, in the *N* gene (59). Moreover an *E. coli* carrying another *rpoB* mutation, *rpoB3595*, supports growth of a λ expressing a defective N (100), a finding consistent with the idea that an interaction with N normally modifies RNA Pol into a termination-resistant form (86).

Host Factors

Mutations in *nusA*, *nusB*, *rpsJ* (*nusE*), and *rho* (*nusD*) that result in a failure of the host *E. coli* to support N-mediated antitermination identified genes whose products were likely to be involved in N action (54). A direct role for the Nus factors in N action was demonstrated in two ways. First, this was shown using an in vitro system with S30 extracts derived from *nus* mutants and controls with plasmids expressing the wild-type allele (34). Extracts from *E. coli* derivatives with *nus* mutations failed to support N antitermination, while extracts from the same mutant bacteria with plasmids expressing the wild-type allele supported N antitermination. Second, a transcription elongation complex containing N and the Nus factors was identified and characterized using an in vitro system that requires S100 extracts (63, 81). The inability to develop an in vitro system with purified components that showed processive N-mediated antitermination indicated that all of the required factors had not been identified. However, N, NusA, and RNA Pol with a template that included a *nut* site was shown to be sufficient to form a minimal transcription complex that can read through a single downstream terminator (179). This led to the idea that NusA and N are the core components of the complex.

These host factors (collectively referred to as Nus) subverted by the phage for N antitermination play important roles for the host in transcription and translation. Subsequent to its identification as a component of the N-modification complex, the *nusA* gene product was shown to influence transcription pausing and termination in *E. coli* (133). The isolation of thermal-sensitive (118) and cold-sensitive (26) alleles of *nusA* identified NusA as an essential protein in *E. coli*.

The *nusB* gene product was shown to influence both transcription termination and antitermination (49, 57, 91, 172). The isolation of *nusB* null mutants that were cold-sensitive defined NusB as being essential at low temperature (166). The *rpsJ* (*nusE*) gene (56, 173), encoding ribosomal protein S10, obviously plays a role in translation. Evidence for involvement of the *rho* gene product derived from experiments with an *E. coli* derivative with the *rho026* allele. This allele was referred to *nusD*, because *E. coli* carrying *rho026* as its only *rho* allele exhibits a slightly weakened ability to support N action (32).

Second-Site Suppressors

Mutations in *E. coli* and λ have been obtained that suppress the effect of *nus* mutations on N action. Certain second-site mutations in the *E. coli nusB* gene as well as mutations in the λ *N* gene and *nut* site enhance the action of N in *E. coli* carrying either the *nusA1* or *nusE71* mutations (53, 150, 171).

The NUT Site (see figure 9-1)

Analysis of a *cis* dominant mutation that blocks the action of N in the p_L operon led Salstrom and Szybalski (148) to identify the *nut* site in the p_L operon. Based on this observation, the *nut* site in the p_R operon was identified (146). Identification of a frameshift mutation in *cro* that allows translation past the normal translation stop in the *cro* gene and thereby blocks N action in the p_R operon provided the first evidence that N action, at least in part, works through the NUT RNA (126).

The NUT site has three elements: BOXA and BOXB, a stem-loop structure, and between them a spacer region (figure 9-2) (52). As will be discussed in detail below, the N and Nus proteins assemble on the NUT site of the nascent RNA and modify RNA Pol to a termination- resistant form. Transcription initiating at p_R is modified by N as it passes through the *nutR* site, causing RNA Pol to transcend the tR1 and *nin* terminators. This allows transcription of the downstream *O* and *P* replication genes, the *nin* region, and the *Q* gene. Similarly transcription from p_L is modified by N as it passes through the *nutL* site, causing RNA Pol to transcend downstream terminators.

Q and qut

The *Q* gene encodes a 20 kDa protein that, like N, modifies RNA Pol to a termination-resistant form that is responsible for transcription of most of the λ genome (78). Unlike N, which recognizes RNA sequences, Q recognizes a DNA sequence called *qut* (159, 160, 185). A more complete discussion of Q can be found below.

Nun

Phage HK022 differs from other members of the lambdoid family in not having an N-like function. Located at the position of the *N* gene, HK022 has the *nun* gene, encoding a protein that acts not with the HK022 RNA but with λ RNA! It was observed that λ fails to grow in an HK022 lysogen even though HK022 grows in a λ lysogen (136). This led to the idea that there is an immunity-independent (immunity describes the operator–repressor interaction) activity in HK022 that blocks λ growth in the HK022 lysogen. This exclusion is specific to phage with the λ NUT site; other lambdoid phages having N-NUT systems that are heterologous to the N-NUT of λ grow in an HK022 lysogen. In contrast to N, which uses NUT sites for antitermination, Nun uses λ NUT sites to terminate transcription in regions of the λ genome that do not contain terminators. Moreover, unlike either intrinsic or Rho-dependent termination, Nun-mediated termination requires the same Nus proteins used in N-mediated antitermination. Thus Nun acts as a highly specific exclusion function, inhibiting, as far as we know, only those lambdoid phages with the *nut* sites of λ.

The New Story

The N Antitermination Protein

Although the N antitermination complex is composed of a number of proteins and an RNA site, in vitro experiments provide evidence that N protein itself is the prime mover in establishing the termination-resistant transcription complex (37, 132). One can eliminate Nus factors (37), and even the NUT site, and efficient antitermination is still maintained (76, 132). However, for such factor- and site-independent action, N must be supplied at very high concentrations and the salt concentration must be low. Under these special conditions the modified polymerase is capable of antitermination of one terminator at a limited distance; but not the processive antitermination through multiple terminators observed only with the full complex. Binding studies show that the affinity of N for the NUT site is approximately 1000-fold higher than it is for an RNA without a NUT site (168). Thus, Nus factors and a NUT site on the RNA are significant contributors under physiological conditions to the efficiency and processivity of N-mediated transcription antitermination.

N and its homologues are, to a first approximation, specific for their cognate NUT sites (47, 150). Functional regions of the N protein were first identified by Lazinski et al. (99) using hybrids formed between the *N* gene of λ and *N* homologs from other lambdoid phages. Their study showed that the NUT site specificity was located in the amino portion of the N protein, an 11 amino acid region rich in Arg-residues. Lazinski et al. (99) pointed out that a number of RNA binding proteins had similar Arg-rich motifs or ARMs. The importance of the Arg residues was confirmed by a mutational analysis in which 14 different

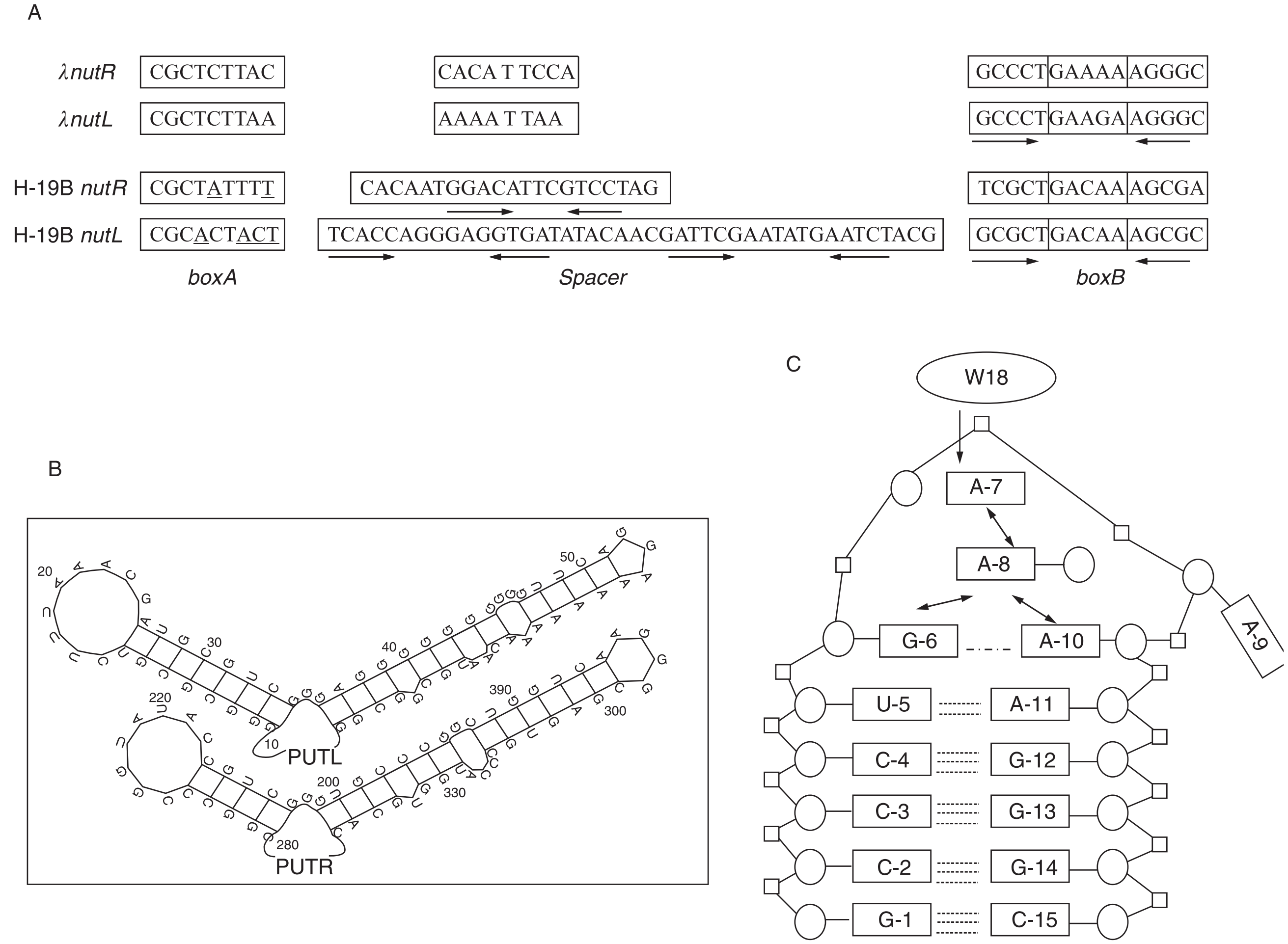

A
λnutR CGCTCTTAC CACA T TCCA GCCCT GAAAA AGGGC
λnutL CGCTCTTAA AAAA T TAA GCCCT GAAGA AGGGC
H-19B nutR CGCTATTTT CACAATGGACATTCGTCCTAG TCGCT GACAA AGCGA
H-19B nutL CGCACTACT TCACCAGGGAGGTGATATACAACGATTCGAATATGAATCTACG GCGCT GACAA AGCGC
boxA
Spacer
boxB
B
PUTL
PUTR
10
20
30
40
50
200
220
280
300
330
390
C
W18
A-7
A-8
A-9
G-6
A-10
U-5
A-11
C-4
G-12
C-3
G-13
C-2
G-14
G-1
C-15

Figure 9-2 Features of antitermination sites in early operons of lambdoid phages. A: Comparison of *nut* regions (L and R) of λ and H-19B with components elements identified below (52, 120). The arrows indicate regions of hyphenated dyad symmetry that are sources of stem-loop structures. B: The PUT stem-loop structures of phage HK022 (6). C: The GNRA fold-like structure formed by λ BOXB upon interaction with N, shown here for the NUTR-BOXB. Bases, sugars, and phosphates are represented respectively by rectangles (with base indicated within), ovals, and squares. Dashed lines shows hydrogen bonds and the –·– line shows sheared G-6 A-10 base pair. Bases are numbered from the 3′ end of the upstream stem to 5′ end of the downstream stem. The purine ring of adenine 7 stacks with the indole ring of Trp18 of N and that guanine 9 is extruded from the structure. See thebacteriophages.org/frames_0090.htm for a color version of this figure.

amino acids in the ARM region of N, including five Arg residues, were systematically replaced with different amino acids (45). Single changes at each of the five Arg residues interfered with N action, while similar changes at six other residues did not affect N action. The nature of the interaction between N and the NUT site is discussed below.

Experiments from the Greenblatt laboratory defined functional roles of various regions of the 107 amino acid residue N protein (114). Affinity chromatography using a set of GST fusions to various fragments of N identified amino acids 34–47 of N as the NusA binding site. A series of experiments that included affinity chromatography, band shifts, and in vitro antitermination assays provided evidence that both regions near the amino (within residues 1–47) and carboxy (including residues past residue 73) termini of N are involved in binding RNA Pol. Despite having apparent functional domains, N appears unstructured in solution, but does become structured when it binds to NUT RNA (114, 167).

Host Factors

(Note: Although the Nus factors have important physiological roles for the host bacterium, we will use the term Nus$^-$ to describe the phenotype of a mutant *nus* allele that is defective in support of N antitermination activity.)

NusA

Genomic, structural, genetic, and functional studies collectively provide an integrated view of NusA action. Orthologs of NusA have been identified throughout the bacterial kingdom. Analysis of these sequences led to identification of three regions that have similarities with known RNA-binding motifs: one S1 (13) and two KH (60). In vivo studies coupled with in vitro studies characterizing *nusA* point mutations showed that the S1 and KH domains are each involved in RNA binding. Moreover, these mutations influenced, to a greater or lesser degree, N-mediated antitermination, and bacterial viability (115, 191, 192). These in vivo studies suggested an order of assembly of the N-Nus complex that is the same as that derived from in vitro studies (113, 179). The assembly begins with N and NusA binding to the NUT site in the newly synthesized NUT RNA, followed by the addition of the other Nus proteins (NusB, ribosomal protein S10, and NusG).

A set of truncated NusA proteins were used in an in vitro study that associated particular regions of NusA with specific functional activities (109): (i) a NusA segment must include the S1 and two KH domains to support N-mediated transcription antitermination, (ii) both the N- and C-terminal regions of NusA are important for RNA Pol binding, (iii) the amino region (amino acids 1–348) is sufficient for enhancing termination at an intrinsic terminator to levels observed with the complete NusA protein (495 amino acids),

The crystal structures of the NusA proteins from *Thermotoga maritima* (0.21 nm resolution) (183) and *Mycobacterium tuberculosis* (0.17 nm resolution) (65) were determined. Even though there are some differences, these structures support the idea of a continuous RNA binding structure that includes the S1 and two KH domains as well as the amino portion of NusA.

NusB

The amino acid change of the commonly used *nusB5* (causes a Nus⁻ phenotype) and the suppressor *nusB101* (suppresses the effects of the *nusA1* and *nusE71* mutations on N action) mutations have been identified; a Tyr to Gly at position 18 (NusB5) and an Asp to Asn at position 118 (NusB101) (25). First identified by its role in N-mediated antitermination, NusB has more recently been characterized for its role in expression of *rrn* operons (154, 161). NusB, in cooperation with NusE, was shown to bind to an RNA containing the consensus BOXA sequence (108, 122) and the NusB and NusE proteins interact even in the absence of RNA (66, 112). The solution structure of NusB as determined by nuclear magnetic resonance was shown to consist of two subdomains, each having an all α helical fold structure (3, 64).

Ribosomal Protein S10 (NusE)

The sequence change of the only identified mutation in *rpsJ* (encodes ribosome protein S10) that affects N-mediated antitermination, *nusE71*, has been identified as an Ala to Asp change at codon 86 (25). Structural studies on the S10 orthologue from *Mycobacterium tuberculosis* indicate little secondary structure (66). Both the NusB and S10 proteins of *M. tuberculosis* (66) and *E. coli* (112) have been reported to form a complex. This could indicate how the two proteins interact in the N antitermination complex. However, there is likely to be another role for S10, since even under conditions where NusB is not required, S10 (NusE) is required for N antitermination (129).

NusG

The original selection for mutations with a Nus⁻ phenotype did not yield mutations in *nusG*. However, Gottesman and coworkers isolated a mutation in *nusG*, originally called U, that, like *nusB101*, suppresses the Nus⁻ phenotype of the *nusA1* and *nusE71* mutations (164). The first evidence indicating that NusG is required for N action came from in vitro studies of Greenblatt and coworkers (103, 104). Although a requirement for NusA, NusB, and S10 in N-mediated antitermination could be demonstrated in vitro, efficient N-mediated antitermination was only observed when the in vitro reaction mixture also included an S100 or S30 extract (34, 63). This suggested that another factor(s) in addition to the identified Nus factors was required for N action. Greenblatt and collaborators identified the missing component as NusG by showing that purified NusG with the previously identified Nus factors eliminated the requirement for the S100 (103). More recently, it has been shown that NusG is also required for N action in vivo (190). Because NusG is essential, bacteria having only a single inactivated *nusG* allele could not be used to assess the role of NusG in N action. To assess physiological activity in the absence of NusG, Gottesman and collaborators developed a system for creating bacteria depleted of NusG (163). Bacteria with the chromosomal *nusG* allele disrupted by insertion of a *kan* cassette (*nusG::kan*) and an intact *nusG* gene carried by a plasmid temperature-sensitive for replication are grown for many rounds of growth at an elevated temperature nonpermissive for replication of the plasmid. This segregates *nusG::kan* derivatives free of the plasmid and, after a number of rounds of growth in the absence of the plasmid, derivatives free of NusG are obtained. *E. coli* depleted for NusG in this way failed to support growth of λ (190). However, λ derivatives that carried either their own *nusG* gene or that were deleted for terminators obviating the need for N antitermination produced phage bursts 30 times greater than did λ infecting *E. coli* depleted for NusG. Thus, the reduced burst of λ in the NusG-depleted bacterium is due to the lack of NusG and a resulting failure in N-mediated antitermination.

In addition to its role in antitermination, NusG facilitates action of termination protein Rho; significant readthrough of some but not all Rho-dependent terminators has been observed in bacteria depleted for NusG (163). A slight decrease in termination at some intrinsic terminators has been reported. However, the main effect of NusG is on Rho-dependent termination (104). Through its direct binding to Rho, NusG was proposed to act as a bridge between Rho and RNA Pol (104, 121). Release of Rho from stalled transcription complexes is slowed by NusG, which only stably binds to the elongation complex in the presence of Rho (121).

The product of the *rho026* allele and the nature of its interaction with NusG is of interest because, as discussed before, *rho026* confers a weak Nus$^-$ phenotype (32). Rho026 differs from wild-type Rho product by not interacting with NusG at high temperature (12). Although the *rho026* mutant shows reduced ability to support N action, it does not exhibit a significant reduction in Rho-dependent termination at any temperature (32). Consistent with these observations, the termination activity of Rho026 is less dependent on NusG than is wild-type Rho (12, 104). Overexpression of NusG reduces wild-type Rho activity presumably by competing with NusG-Rho complexes for RNA Pol binding (12). NusG enhancement of Rho-dependent termination appears to be specific to the *E. coli* RNA Pol. Although Rho-dependent termination can be observed with T7 and Sp6 RNA Pol, the termination is not enhanced by NusG (128).

The NUT Site

Information gained from genetic and structural studies as well as genome comparisons has extended our understanding of the nature and function of interactions at NUT RNA. The *nut* sites have been divided into three components: *boxA*, a spacer, and *boxB* (figure 9-2A). Although the interactions at NUT involve a complex formed between all of the components, the following discussion will focus on each of the component sites individually.

BoxA

A comparison of *boxA* sites yielded a consensus sequence, referred to as *boxAcon* (CGCTCTTTA) (55). The *nut* sites of most lambdoid phages, with the exception of *Salmonella* phage P22 which has the consensus sequence, have slight variations from the consensus sequence (e.g., that of λ is CGCTCTTac). Phages with *nut* sites having the consensus *boxA* sequence require a less effective N-Nus complex than do phages having *nut* sites with variations of the consensus *boxA* sequence (55). The consensus *boxA* suppresses the effect of *boxB* mutants that cause a greater than 1000-fold reduction in affinity for N (29). Because the consensus BOXA binds NusB (and S10) more avidly than does the λ wild-type BOXA (108, 122), it is likely that more secure binding of NusB stabilizes weak binding of NusA with N in the absence of BOXB.

BOXA was studied using a reporter construct whose relevant features include a *galK* reporter gene fused to the *lac* promoter with an intervening cassette containing the λ *nutR* and the downstream tR1 Rho-dependent terminator (129). N was provided in trans and *galK* expression provided a measure of the effectiveness of antitermination. Although the *boxA16* (CGCTaTTAC) (145) and *boxA5* (CtCTCTTAC) (127) point mutations reduced N-mediated antitermination 35- and 7-fold respectively, a deletion that included *boxA* and upstream sequences was not defective for N-mediated antitermination. Surprisingly, the deletion suppressed the defect in N action normally imposed by the *nusB5* mutation, but not the block imposed by mutations in other *nus* genes. Thus, NusB activity is required when *boxA* is present and dispensable when *boxA* is deleted. Experiments with the *nusB5* mutant provided further insight into this apparent paradoxical finding.

The *nusB5* host, which expresses low levels of an active NusB (25, 112) does not support λ growth due to a failure in N action. However, when NUT RNA with the BOXA5 change is overexpressed from a plasmid in the *nusB5* host, λ growth and, presumably, N action, are supported. Further, when λ NUT RNA with the wild-type BOXA is similarly overexpressed, λ growth is not supported.

These results are consistent with a model in which there is an inhibitory factor that not only binds BOXA RNA but also binds more avidly to BOXA5 RNA. Presumably, it is this preference for the inhibitor that causes the *boxA5* mutation to reduce the effectiveness of N action in the host with wild-type *nus* alleles.

The Inhibitor

In brief then, the BOXA site within the NUT RNA recognizes an inhibitor that interferes with the formation of an effective N antitermination complex. It was further proposed that NusB, by binding to BOXA, hinders access of the inhibitor to the NUT site and thereby prevents the action of the inhibitor (129). Results of an in vitro experiment showing that "substantive" levels of NusB were not necessary for N-mediated antitermination supported this model. It was argued that NusB is not necessary in vitro because the free inhibitor is not present (37).

There is no direct evidence as to the nature of the putative inhibitor. However, there is suggestive evidence that the C-terminal domain of the α subunit (the αCTD) of RNA Pol plays a role in the action of the inhibitor. A point mutation in *rpoA* has been identified that interferes with N action (125), while two other point mutations (D305E and L280H) that alter the αCTD as well as a deletion (*rpoA*△*3'*) that encodes an α subunit missing the CTD were shown to enhance N antitermination (151). The fact that a CTD-truncated α enhances N antitermination suggests that, if left unchecked, the αCTD can negatively impact on N action; i.e., acts to inhibit N-mediated antitermination (50). In vitro studies by Hanna and colleagues on interactions of α during transcription elongation are consistent with these ideas. The αCTD was shown to interact with the nascent RNA during transcription—an interaction that is not observed in the presence of NusA (105). In vitro, as in vivo, N antitermination is effective even when the α subunit is missing the CTD (106). Moreover, it is unlikely that the αCTD itself is the inhibitor. In vitro studies using purified components to study N antitermination have RNA Pol with the complete α subunit. If the αCTD is the inhibitor, then NusB should be required in these in vitro reactions to block the inhibitory activity of the αCTD. However, NusB has little effect on N action in this type of in vitro antitermination reaction, unlike the other Nus factors that are required (37). Further, it is unlikely that free α is the elusive inhibitor since the mutant αCTDs enhance N activity in merodiploids in which a wild-type α is also expressed (151). What we do suggest is that the inhibitor, in some way, works through the αCTD and NusB supports N antitermination by blocking action of the competitor that acts through the αCTD.

BoxB

The specificity of the different N proteins for their cognate NUT site indicated that sites of N recognition are most likely located in the regions of the *nut* sequences showing significant sequence heterogeneity. Comparison of *nut* sites from different lambdoid phages identified sequences composing the BOXB stem-loop structures (44, 99) as having the heterogeneity that would allow for binding specificity. The first mutation in a *nut* sequence that eliminated N action was located in the loop of the *nutL boxB* (148). DNA oligonucleotides complementary to BOXB RNA severely inhibited N-mediated antitermination, while DNA olgonucleotides complementary to sequences immediately upstream and downstream of the NUT site had little effect (180).

In an elegant set of experiments, the Das group (99) used hybrid *nut* sites and chimeric N proteins constructed from different lambdoid phages to conclusively demonstrate that BOXB is the site that provides specificity for N and that the amino portion of the N protein recognizes the NUT RNA sequence. The following provides examples of the experiments that led to these conclusions. An antiterminator tester plasmid carrying a hybrid *nut* site with the *boxB* from phage 21 and the *boxA* and spacer from λ was far more effective in supporting antitermination when supplied with the N of phage 21 than when supplied with the N of λ. Further, a hybrid N protein with the first 35 amino acids from phage 21 and the C-terminal 74 amino acids from λ is active with NUT21 but not NUTλ. Based on the observation that the amino proximal regions of all N proteins share a motif rich in Arg residues, Lazinski et al. (99) proposed that this motif is involved in binding of N to RNA. Further, they suggested that ARMs are important in binding of other proteins such as the human immunodeficiency virus TAT protein to RNA. Subsequent studies bore this prediction out (165). The ARM of N is functionally important since, if the Arg residues are changed, N action is impaired (45).

Nuclease protection, mutational analysis, and band shifts provided important information as to the nature of N and NusA binding with the BOXB stem-loop (18, 21, 38, 115, 167). Both the stem-loop structure and nucleotides in the ascending arm and the loop are essential for N binding to BOXB RNA. Not all mutations in the loop that reduce effectiveness of the NUT sequence affected N binding (18).

Structural studies have addressed the question of the nature of N interaction with BOXB (101, 114, 149, 162, 167) and the following briefly summarizes some of the important information gained from those studies. NMR studies with peptides containing the ARM of λ N have probed the nature of binding between BOXB and N. These studies revealed that the ARM domain of N assumes a bent α-helical structure and the BOXB loop assumes a structure resembling a GNRA fold, a tetranucleotide structure that stabilizes the RNA hairpin (4, 79). Consistent with such a structure, the bound BOXB loop includes a sheared GA pair (20), side by side pairing of the GA pair that is seen in GNRA folds. Genetic and structural studies provide evidence that the BOXB GNRA-like fold is necessary for N binding. Further, the purine ring of adenine 7 in the loop stacks with the indole ring of Trp 18 of N and base guanine 9 in NUTL and, by inference, adenine 9 in NUTR are extruded from the structure (see figure 9-2C for numbering of bases in the BOXB stem-loop). The BOXB structure differs from the GNRA tetraloop, because proper folding of BOXB requires peptide binding with stacking between the indole ring of Trp 18 of N and base A-7 of BoxB. Although a change at position 9 does not affect N binding, it affects biological activity (38) and can eliminate NusA binding (115). Thus, the structure of BOXB appears to be important for binding both N and NusA. Studies with P22 N (80, 102) show that the P22 BOXB with bound P22 N (also called 24) is in a structure resembling a GNRA fold and the ARM of the bound N also assumes a bent α-helical structure (14). However, in this case a different base is extruded from the loop sequence. The difference in the extruded base, which in the case of λ N appears to interact with NusA, could explain the differences in requirements for NusA exhibited by the Ns of λ and P22 (27, 49).

Cooperative Interactions at NUT

We will now examine more carefully the integration of the individual interactions that result in the generation of the full complex. A stable N antitermination complex forms only after transcription passes through the *nut* site. This complex contains N and the Nus factors (8, 31, 81).

Mobility shifts were used to assess binding of the other *E. coli* factors to a core complex consisting of NusA, N, and RNA Pol, with λ NUT RNA (116) . When the RNA had a wild-type λ BOXA sequence, binding of NusB, S10, and NusG was found to be cooperative; only when all three proteins were present was the core complex effectively shifted. Most changes in BOXA away from consensus reduced this binding to the core complex. However, when the NUT site had BOXACON (the consensus BOXA sequence), NusB and S10 were sufficient to shift the core complex.

These results appear to contradict previously discussed in vivo experiments showing that, when BOXA and some upstream sequences were deleted, S10 was required for N action even though NusB was not (129). However, the interaction of S10 in the absence of NusB and/or a proper BOXA sequence with the core complex may not be observable by gel shift assay.

Employing a technique called an immunoprint, anti-N antibody was used to precipitate transcription complexes with labeled nascent RNA (8, 31). The in vitro transcription reaction included N protein, an S30 extract, and a DNA template with a *nut* site. Transcription was stopped throughout the length of the template using chain-terminating nucleotides. The RNAs isolated from the precipitated complexes were separated by gel electrophoresis and the results could be used to determine when N became associated with the transcription complex. Controls showed that precipitable complexes were not formed if a purified NUT containing RNA was mixed with RNA Pol and the other components of the reaction; i.e., the anti-N antiserum precipitated only components in an elongation complex.

One set of experiments using the immunoprint technology showed that N association with the transcription complex is observed beginning with those transcripts that have reached the distal end of the *nut* site (the 3′ end of *boxB*) and those that have progressed beyond that point (8). In these experiments, N and the S30 extract were present during the entire reaction. This result is consistent with the NUT RNA directing acquisition of the N-Nus complex.

A second type of experiment using the immunoprint technology studied acquisition of N by paused transcription elongation complexes preformed in the absence of N and S30 extract (8). N and S30 extract were added to paused RNA polymerases that had been transcribing the template containing a *nut* site. Only one stable immune complex was precipitated with anti-N antibody; that transcription complex had proceeded just four nucleotides beyond the 3′ end of the BOXB stem-loop. Moreover, stable complexes that are precipitable with immune serum did not form with transcribing RNA Pol that had passed the *nut* site. It was only during the limited time when the BOXB structure was still forming that N could be acquired. This provides strong evidence that the antitermination complex with N assembles only when the RNA Pol is at the *nut* site.

Under completely different experimental conditions using a purified transcription system, it was shown that RNA Pol can be modified to antiterminate even when NusA and N are supplied after transcription has advanced a significant distance past the *nut* site (180). That is, the NUT RNA can apparently bind N and NusA and loop over to attach to the distally located RNA Pol. Although the latter results are intriguing, we find the data with S30 extracts more compelling because the conditions of these experiments are more representative of what happens in vivo.

The results of the immunoprint experiments with the S30 extracts suggest that the initial interaction of N with NUT RNA may be very different than the one proposed by studies of N binding to purified NUT RNA. We interpret these S30 results as showing that N binds to the sequences of the ascending stem and loop of BOXB that have advanced out of the RNA–DNA hybrid in the absence of formation of the stem-loop structure. When transcription reaches four bases beyond BoxB and Nut, the RNA pol–DNA–RNA ternary complex is presumably in its native transcription elongation state. In this state, the 3′ end of the RNA is within the active center and the upstream eight nucleotides are in a hybrid with the template DNA (96, 155, 184). An additional six nucleotides upstream of the RNA/DNA hybrid is protected from nuclease as it passes through the RNA Pol to the exterior (95). At this point, the upstream part of NUT, which includes BOXA, will be completely extruded outside of the RNA Pol and thus should be able to bind NusB and S10 (NusE). The BOXB stem, however, cannot yet be formed, as most of the downstream side of the stem will still be trapped in the RNA/DNA hybrid. How can N bind to the complex if, as suggested by a number of biochemical and structural studies with pure components, N binding requires that the BOXB stem with its tetraloop must form?

Because BOXB is within the RNA Pol complex at this +4 position, it is possible that this association alters the RNA so that it binds N even though the BOXB stem-loop has not formed. According to this scenario, formation of the BOXB stem-loop would be unnecessary for the initial binding of N. As the RNA is extended, the stem sequences would be released from their tight association with RNA Pol and the BOXB stem-loop could form. Retention of N, and thus maintenance of the antitermination complex, would then require that N bind to this newly formed structure. Processive antitermination would proceed with N binding to the BOXB stem-loop and association of the other Nus factors through binding at BOXA and by protein–protein interactions.

To explain the rapid assembly of the antitermination complex at the NUT site, we consider N action in light of an idea originally proposed by Das and collaborators (33). As they suggested, N and NusA are bound in a loose association with RNA Pol prior to reaching the *nut* site. When NUT RNA is synthesized, N can rapidly slide into position on the BOXB sequences comprising the leading arm of the BOXB stem and the loop region. This can occur prior to the formation of the BOXB structure, while sequences composing the descending stem are still in the RNA–DNA hybrid. As discussed above, this is apparently the one position during transcription when NUT RNA can bind N and Nus factors to initiate formation of a stable RNA–N-Nus–RNA Pol complex. Hence, the role of BOXB may not be to increase the local concentration of N to facilitate binding to RNA Pol but to allow N to form a stable complex. RNA Pol with NUT RNA and N and Nus proteins then becomes a processive transcription complex (34, 113).

N Action

How does N-modified RNA Pol become antitermination-proficient for both Rho-dependent and intrinsic terminators? At intrinsic terminators N could shift from the NUT stem-loop temporarily to one arm of the terminator to transiently block formation of the terminator stem-loop (76). We suggest that at Rho terminators, the properly positioned NusA could block access of Rho to the DNA–RNA hybrid. This would prevent Rho from destabilizing the hybrid. For alternatives on the possible mechanism of N, the reader is referred to other reviews (33, 73, 133). These models primarily address the question of antitermination at intrinsic terminators. Therefore, we will look in more detail at N-mediated antitermination at Rho-dependent terminators.

We view NusA as likely being the critical element that prevents Rho-dependent termination. In addition to its action in N and Q antitermination systems, NusA plays a role in antitermination in the *rrn* operons (169). All these antitermination systems support readthrough of Rho-dependent termination sites. Importantly, NusA can have a direct effect on Rho termination. In a pure transcription system with only DNA, RNA Pol, and Rho present, termination is reduced by NusA at Rho-dependent terminators (11, 98, 156).

One key observation for understanding the role of NusA in promoting readthrough is that in an in vitro transcription system NusA prevents termination at the first (site I) of three subsites within the Rho-dependent terminator tR1. However, NusA does not prevent Rho-dependent termination at the downstream sites II and III in tR1 (98). Above we argue that NusA occupies a weak binding position on RNA Pol, and it is this binding that we propose interferes with the initial interaction of Rho with RNA Pol.

We suggest that Rho displaces NusA from RNA Pol at the first subsite but fails to terminate transcription. However, Rho is now in a position to terminate transcription at all of the downstream sites. Studies with the *nusA11*(Ts) mutation (118) provide an insight into how NusA and Rho may interact. This mutation causes a Rho-defective phenotype with characteristics similar to the *rhots15* mutation (118; D. L. Court, unpublished data). NusA11 appears then to act as an antagonist of Rho activity, possibly by binding more tightly to RNA Pol than does wild-type NusA, and thereby becoming more resistant to Rho-imposed dissociation from RNA Pol. Identification of a second-site mutation in *rpoC* (encodes the β′ subunit of RNA Pol) that suppresses the Rho-defective and temperature-sensitive phenotypes of *nusA11* is consistent with this idea (85). The altered β′ subunit could reduce the binding of NusA11 to RNA Pol.

NusA has been postulated to modulate Rho activity in another way. By stimulating pausing of RNA Pol, NusA could facilitate coupling of transcription and translation (16, 17, 42, 147, 189). The length of the gap between the transcribing RNA Pol and the trailing ribosomes is thought to influence Rho action; i.e., the larger the gap, the larger the target for Rho action (1). By binding tightly to RNA Pol, NusA11 could increase pausing of RNA Pol. This would allow the translating ribosomes to keep up with the transcribing RNA Pol; i.e., stronger coupling of transcription and translation and, in turn, reduced Rho activity.

Thus, NusA could reduce Rho termination by interacting with RNA Pol in two independent ways. How does this relate to the antitermination activity of N? We suggest that merely by facilitating the interaction of NusA with RNA Pol, N interferes with Rho action.

Variation of the NUT Site

H-19B (158), a lambdoid phage (82) carrying the gene encoding a Shiga toxin, *stx*, differs from the λ paradigm in both its *nut* site and host requirements (119, 120). H-19B shares the same N-NUT system with two other identified lambdoid phages: 933W (130) (another phage carrying an *stx* gene) and HK97 (88). This type of *nut* site (figure 9-2A) has two major differences from the λ paradigm (120). The first difference is in the *boxA* sequence, which is very degenerate compared with those of the λ type. Unlike λ, H-19B grows well in *E. coli* with either or both the *nusB5* (or a *nusB::cam* insertion mutation) and *nusE71* mutations (120). Hence, this phage, which does not appear to have a functional *boxA* sequence of the λ type, does not appear to require NusB or S10 (NusE), two proteins known to bind to wild-type BOXB (55, 108, 122, 129).

The second unusual feature of the H-19B type of *nut* sites is the presence of hyphenated dyad symmetries in the spacer regions (figure 9-2A): two in *nutL* and one in *nutR* (120). Mutational analysis of *nutR* showed that nucleotide substitutions weakening formation of the spacer stem structure interfere with NUT activity. However, compensating mutations that change the sequence but allow stem structure restore NUT function. Interestingly, sequences composing the entire stem-loop can be deleted without a significant loss of function. It was concluded that the stem structure serves to reduce the linear distance along the NUT RNA, in this way serving as a “reducer” by bringing separated regions of the NUT sequence together into a functional unit (120). The sequences in the “reducer” stem-loop of 933W and

HK97 are the same as those in H-19B. This evolutionary conservation suggests that there may be an additional, as yet unidentified, role for the reducer stem-loop.

Translation Regulation of N Expression

The importance of N as a regulatory protein is reflected in the number of ways the expression of the *N* gene is controlled. As the first gene of the p_L operon, the *N* gene is regulated at the transcription level by CI and Cro repression p_L (131). At the post-translation level, the Lon protease degrades N protein causing a relatively short protein half-life (68). Studies with λ have identified two other regulatory mechanisms that modulate N expression, at the level of translation, through sites encoded in sequences in the 223 base long *N*-leader (figure 9-3). The first of these sequences is the *nutL* site and the second is a stem-loop structure sensitive to the double-stranded RNA-dependent endonuclease, RNaseIII (24). This site, the N ribonuclease III (three) site, is referred to as N(RTS) (35).

The following discussion of these regulatory mechanisms is based primarily on three papers — Wilson et al. (181), Kameyama et al. (90), and Wilson et al. (182) — and is illustrated in figures 9-3 and 9-4. It was surprising to discover that the NUT site, which we have seen is a key element in the action of N in modifying the transcription complex, is also involved in controlling N synthesis at the level of translation. Even more confounding was the discovery that the very Nus products required for N's transcription antitermination activity are also required for this translation repression. This regulatory process relies on a complicated interplay between the N(RTS), the NUT site, and the N ribosomal binding site. The RNaseIII-sensitive stem-loop structure can be viewed as being a central player in translation control. Located just prior to the *N* gene this large stem-loop structure in the RNA has two roles: first, it acts as a direct translation inhibitor, preventing ribosomes from easily binding at the *N* initiation region; second, it holds the *N* gene initiation codon in the correct position relative to NUT for autorepression when N and the Nus factors bind to the NUT site. The stem is a substrate for RNaseIII (107), and its cleavage by RNaseIII prevents both types of repressive effects on N synthesis. Hence, translation repression is best observed in an RNaseIII mutant. In the absence of cleavage of the N(RTS), the newly expressed N protein acts by binding at the NUT site to block access to the *N* ribosome-binding site. This activity of N, like the antitermination activity, requires the participation of the Nus factors. Translation repression is specific for expression of *N* since translation of downstream genes is not affected. Cleavage by RNaseIII separates this ribosome-binding site from the NUT site and thus eliminates N repression. An *N* leader deleted for sequences encoding the entire N(RTS) maintains N-mediated translation repression. This shows that the stem structure itself is not essential for translation repression by N. RNaseIII does not prevent translation repression when the N(RTS) is not present because the processing site has been removed. Hence, RNaseIII's role in translation repression is due to its action at the *N* leader site.

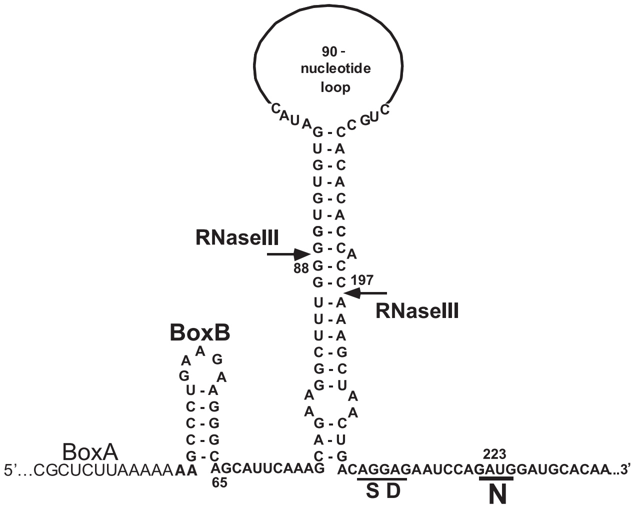
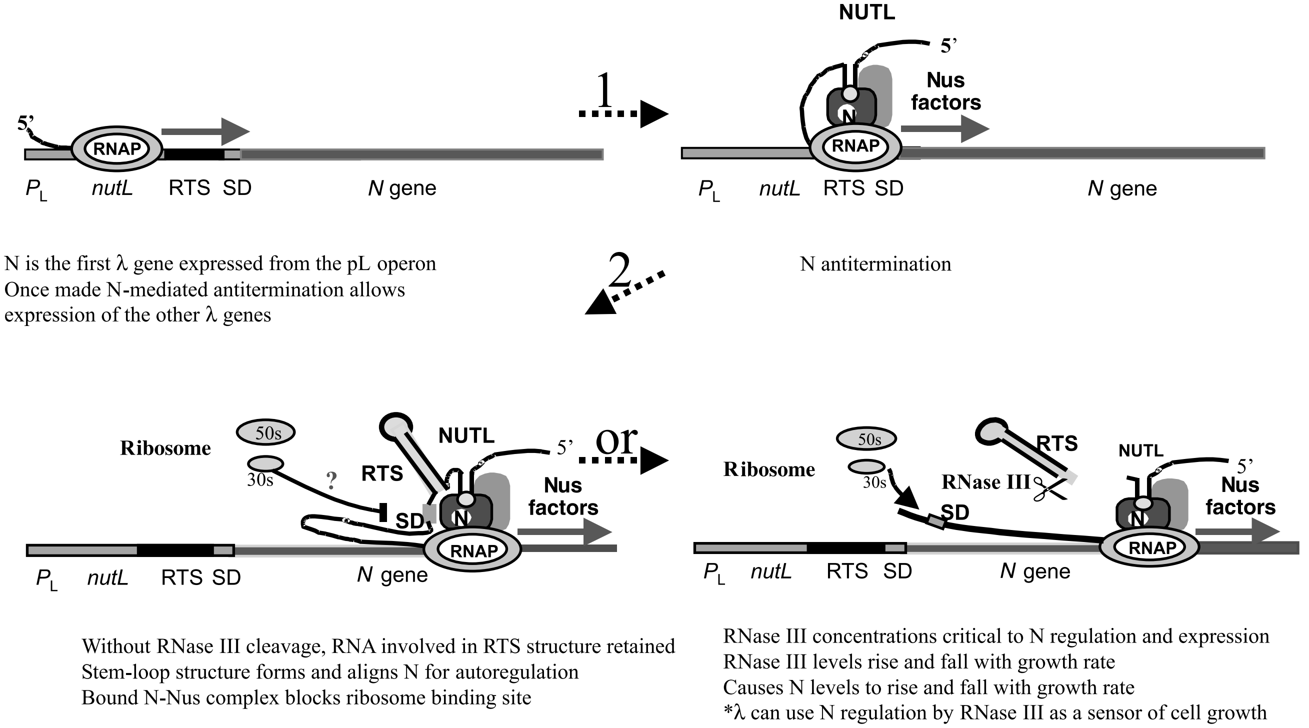

90 - nucleotide loop
RNaseIII
88
197
RNaseIII
BoxB
BoxA
5'…CGCUCUUAAAAAAA AGCAUUCAAAG ACAGGAGAAUCCAGAUGGAUGCACAA..3'
65
223
S D
N

Figure 9-3 The *N*-leader transcript beyond P_L.z The sequence is shown starting at the BOXA sequence of NUTL; numbers indicate distance from RNA start of P_L. The structure of the RNaseIII site is shown with the position of cleavage sites marked by arrows. The N ribosome-binding site is underlined.

In summary, the N(RTS) and RNase III action at that site regulate N expression by influencing translation initiation. The structure per se partially interferes with ribosome binding and N-mediated repression provides an additional block by sterically inhibiting ribosome binding. RNaseIII and *N* expression both increase with increasing growth rate (182) consistent with the hypothesis that RNaseIII activity stimulates *N* gene translation through cleavage of the inhibitory hairpin in a growth-rate-dependent manner. The relationship between RNaseIII activity and *N* expression can thus tie lambda development to the physiology of the host cell.

The HK022 PUT Sequence

Phage HK022 has PUT sequences downstream of the p_L and p_R promoters in place of the NUT sequences found in other lambdoid phages. Although other phages with PUT sites have recently been identified (R. Farooque and R. King, personal communication), all the work to date on this antitermination system has focused on HK022. Unlike NUT, PUT appears to modify RNA Pol to a termination-resistant form without the aid of either phage- or host-encoded proteins (22, 92, 124). The *put* sites (figure 9-2B) are composed of two regions of hyphenated dyad symmetry separated by one nucleotide (6, 22). Studies with mutant *put* sites show that PUT RNA must form both stem-loop structures to be



N is the first λ gene expressed from the pL operon
Once made N-mediated antitermination allows
expression of the other λ genes

N antitermination

Without RNase III cleavage, RNA involved in RTS structure retained
Stem-loop structure forms and aligns N for autoregulation
Bound N-Nus complex blocks ribosome binding site

RNase III concentrations critical to N regulation and expression
RNase III levels rise and fall with growth rate
Causes N levels to rise and fall with growth rate
*λ can use N regulation by RNase III as a sensor of cell growth

Figure 9.4 Assembly of transcription antitermination complex. N translation repression modes. RNA polymerase (RNAP) is shown transcribing from the P_L promoter through *nutL* and the RNase III site (RTS) into the *N* gene. The antitermination complex is formed first at NUTL (to the right of step 1) and the RTS is transcribed and assembles beyond step 2. If RNaseIII is not immediately present the Shine–Dalgarno (SD) is protected in some way from ribosome binding. If RNaseIII is present then the RTS is processed and this allows optimal translation of N. Note that the antitermination complex may remain bound to RNA polymerase. See thebacteriophages.org/frames_0090.htm for a color version of this figure.

active (92). Moreover, to be effective, the PUT site must remain tethered to RNA Pol (153).

The *nusA1* mutation does not influence PUT activity (124) and the search for host mutants uncovered mutations only in sequences encoding the amino proximal domain of the β′ subunit of RNA Pol (22). These changes occur in a region with a zinc finger motif (having a group of four invariant cysteine residues) and did not confer any other observable phenotype. Nuclease protection experiments on stalled transcription complexes at *put* showed that the central part of the downstream stem and the 3′ arm of the upstream stem were specifically protected (153). Moreover, once the RNA Pol passed through the *put* site, the DNA that had already been transcribed was shown to no longer be necessary for downstream transcription antitermination. An *rpoC* RNA Pol mutant defective for PUT-directed antitermination failed to show significant protection of PUT RNA.

Based on the idea that the large number of basic residues in the zinc finger may serve as a surface that interacts with RNA, Sen et al. (152) changed each of the basic residues in the zinc finger to Ala. Although none of these changes was lethal to the bacterium, a number of them failed to support growth of HK022. Many of the mutants functioned better with *putL* than with *putR*, leading Sen et al. (152) to propose that the zinc finger distinguishes between the stem-loop of the two PUT RNAs. Because some of the mutant *rpoC* products were more defective in vitro than in vivo for the same PUT action at the same terminators, it was suggested that there may be additional factors active in vivo supporting PUT-mediated antitermination despite the genetic evidence to the contrary.

The HK022 Nun Protein

Although not identified as a regulatory protein for its own genome, the action of the HK022 *nun* gene product is so tied to the action of λ N that it warrants discussion in this review. The HK022 *nun* gene is located in the same relative position on the HK022 genome as the various *N* genes are located on the genomes of other lambdoid phages. However, the Nun protein does not act as an antitermination protein for HK022 but, instead, acts by terminating transcription on the λ genome at sites not normally involved in termination (reviewed in: 67, 177, 178). Expressed by the HK022 prophage from a constitutive promoter immediately upstream of the *nun* gene (93), Nun acts specifically with NUT sites of the λ type, but not with NUT sites differing from that of the λ type (136). Acting through the λ NUT sites, Nun arrests transcription initiating from the early λ promoters (83). As shown for transcription initiating at the λ p_L promoter, Nun action generates short RNAs spanning 100 nucleotides downstream of *nutL* (157). Action of Nun, like action of λ N, requires the four Nus factors (83, 144). Consistent with these requirements, Nun competes with λ N for binding to the λ BOXB RNA (19, 83). As observed with N, the amino portion of Nun is rich in Arg residues (ARM) and binds to the BOXB structure in a manner similar to the binding of N (19, 41, 77). However, in vitro Nun arrests transcription and blocks the translocation of RNA Pol without causing its release, and in vivo terminates transcription, causing dissociation of the ternary complex (83, 84). In vivo studies with hybrid proteins expressed from hybrid *N-nun* genes provide evidence that the nature of binding to BOXB does not determine the mode of action of Nun or N. For example, a hybrid N-NUN protein having the amino one third from NUN and the carboxy two thirds from N acts at the λ NUT site to promote transcription antitermination (77).

The folding of the wild-type Nun protein inhibits its own binding to RNA, an inhibition that is relieved by Nun binding of NusA. The self-inhibition, which is caused by the C-terminus folding back on the ARM, is particularly noticeable in the presence of Zn^{2+} (174).

Trp 108, the penultimate amino acid, is essential for Nun action. A transcribing elongation complex including the nascent RNA that is modified by a mutant Nun protein missing this Trp residue can freely switch from one DNA template to another DNA template, an action not permitted by the wild-type Nun protein. It was proposed that Nun interferes with transcription elongation by a braking mechanism; the Trp residue on the Nun protein bound to RNA Pol intercalates into the minor groove of the DNA (176). A crosslinking experiment confirmed that, at sites of Nun-mediated termination, the ultimate amino acid on the Nun associated with the stalled RNA Pol is linked to the DNA template (175).

Recent studies suggest that Nun, like N, represses the translation of the λ *N* gene through a complex formed at NUTL (M. Gottesman and D. Court, unpublished result). This provides confirmatory evidence that highly similar mechanisms are used by both N and Nun to carry out antitermination and termination, respectively, using the NUT site. In fact, under very special conditions, Nun, acting through a mutant NUT site, does not terminate λ transcription but, in a limited way, suppresses transcription termination (145).

The Q Transcription Antiterminator

The *Q* gene of λ, located 5.8 kb beyond the p_R promoter and distal to at least four transcription terminators, requires N antitermination for expression. Q function, in turn, is required for full expression of the phage late genes (39, 87) and serves as a second antitermination factor (figure 9-1) (43, 139). *Q*, like N, can antiterminate Rho-dependent and Rho-independent terminators (58,186). As we will see the requirements for and the mechanics of modification of the RNA Pol complex are very different for Q and N.

Q, like N, has a specific binding site, *qut* (*Q utilization*), which is associated with the strong $p_{R'}$ late promoter (figure 9-1) (159). However, unlike N, which works with λ and non-λ promoters provided they are upstream of the *nut* site (36), Q binds to the *qut* DNA sequence and, because *qut* is an integral part of the $p_{R'}$ promoter, only antiterminates transcription initiating at $p_{R'}$ (143, 187). Unlike N, which binds RNA and interacts with a large complex of host Nus factors, Q stabilizes polymerases initiating from the $p_{R'}$ promoter independently of host factors. However, NusA has been shown to stimulate antitermination activity of the λ Q from 3- to 10-fold in vitro (69), but only marginally influences antitermination activity of the Q of another lambdoid phage, 82 (186). Once Q has modified RNA polymerase at $p_{R'}$, the antitermination complex continues through the Rho-independent $t_{R'}$ termination site 196 bases downstream and ultimately through the entire late lytic gene operon, which is 26.5 kb long, with very little drop-off in transcription level. It is not known whether any transcription terminators exist beyond $t_{R'}$ in this long segment. Nor is the terminator known that stops Q-dependent transcription beyond the late genes. N antitermination can also proceed over long distances as demonstrated by transcription from the p_L promoter in the prophage extending all the way to the bacterial *gal* operon more than 20 kb away (2). The level of N antiterminated transcripts that reach the *gal* operon is greatly reduced relative to those that began at the p_L promoter. Between p_L and *gal* there are many terminators, and it has been shown that N does not completely overcome them. However, it is not known whether Q antitermination is highly processive or whether there are just very few terminators in the late operon. In any case, transcription over the entire length is well maintained.

Although Q activity is more efficient in cis, Q can act in trans (40). For example, an infecting λ Q^+ phage can activate antitermination and late gene expression from the $p_{R'}$ promoter-*qut* site of a repressed λ*imm*434 prophage, a phage with a different immunity region but the same $p_{R'}$-*qut* as λ (28). Note that in the lysogen the $p_{R'}$ promoter is constitutively active and is primed for Q action (140). All identified lambdoid phages have Q-like functions. Some, like those of P22 and λ, are nearly identical and functionally interchangeable (142): a Q from one will activate antitermination from the *qut* site of the other. On the other hand, the Q proteins of phages 82 and 21 vary greatly in sequence from each other (as well as from the Q of λ) and only activate their own *qut* sites (186).

Roberts and collaborators have provided the bulk of our knowledge about how Q protein causes antitermination. Their purified in vitro transcription studies have answered many of the initial questions. Early on it had been determined in vitro that λ Q protein and the λ $p_{R'}$ promoter region as well as the first 20 or so bases downstream of the $p_{R'}$ start are required to form a transcription antitermination complex with RNA Pol holoenzyme (159, 185). Q has a limited window during the transcription process to modify polymerase; it does not act during the transcription initiation event at $p_{R'}$ nor does it alter the polymerase elongation complex once the complex is well away from the promoter (69). This window appears to be the brief transition time between the initiation stage and the elongation stage of RNA transcription (143). The details of these initial requirements have now been worked out and a more complete picture of the process is available.

RNA Pol initiating transcription at $p_{R'}$ pauses after synthesizing a transcript of 16 nucleotides (69, 137). At that point, polymerase elongation is delayed by a specific binding of its σ^{70} component (75) to the nontemplate strand of the DNA bubble (94, 134, 135). This binding occurs downstream from the promoter start at a site that resembles a −10 promoter sequence. Thus, σ^{70} rebinds to the DNA after initiation but before it can be released from the stable elongating RNA Pol complex. This rebinding causes a 30 second pause in vitro and a 25 second pause in vivo during a transcription without Q protein present.

Marr et al. (110) examined the nature of the paused RNA Pol. They found that the paused complex with σ^{70} is very similar in structure to the open complex of RNA Pol found at promoters. It differs, however, in having a 16 nt RNA product partially hybridized to the template strand within the transcription bubble. Q binds to this paused intermediate making contact with the DNA upstream of the transcription bubble. In this elongation type of "open complex," σ^{70} contacts the DNA slightly differently from the way it does in the open complex found at the promoter (110, 135). Region 4 of σ, the segment of σ that contacts the −35 segment of the promoter, contacts Q. Region 2, the segment of σ that binds the −10 region of promoters by contacting the unpaired nontemplate strand in the open complex, binds the nontemplate strand at the pause site in much the same way as at the promoter. Region 3 of σ, however, is not properly positioned relative to the rest of σ when compared with its position on the open complex at the promoter. This probably is caused by the presence of the RNA in the complex. When Q is present, it displaces region 4 of σ and binds the DNA segment flanking what would be the −35 position of the paused "open complex." This position, of course, is moved downstream relative to the $p_{R'}$ promoter and is located between what were the −35 and −10 sequences of the normal promoter. Q binds to the DNA as a dimer (110, 188) activating the elongation complex for antitermination. As RNA Pol moves away from the promoter, σ^{70} is dissociated and Q joins the RNA Pol complex.

The specificity and interactions of Q and σ in the paused complex have been confirmed by genetic studies. Mutations in the *qut* site at positions 15 and 13 bases upstream of the normal transcription start site prevent Q binding but do not prevent pausing at +16. These mutations define the Q binding site on the DNA. Mutations in the downstream σ binding site, at positions 2 and 6 bases beyond the normal transcription start, prevent renewed σ binding and pausing of polymerase at +16. All these mutations prevent antitermination, demonstrating the functional requirement for both Q and σ binding (89, 186, 187). Using the +2 and +6 mutations that prevent σ rebinding, RNA Pol can be artificially paused at the +16 position of the mutants by withdrawing appropriate nucleotide substrates. This artificially paused polymerase cannot bind Q and fails to antiterminate transcription, demonstrating the special need for σ binding and a likely interaction of σ with Q (135, 187).

Although only RNA Pol holoenzyme initiating from the $p_{R'}$ promoter and its associated *qut* site are required for Q antitermination in vitro, other components stimulate Q antitermination as judged by in vitro and in vivo experiments. When polymerase is trapped at the pause site by σ binding to the nontemplate strand, it becomes constrained and tends to slip backward into an arrested complex in which the 3′ end of the RNA is no longer at the active center of polymerase. Q does not bind or modify this back-tracked form of polymerase. Two factors associated with RNA Pol, GreA and GreB, cause polymerase to cleave the RNA at the active center in a back- tracked state (9) allowing polymerase to resynthesize to the +16 pause and thereby providing Q with another chance. Roberts and coworkers have shown this stimulation by GreA and GreB using both in vitro and in vivo studies. *E. coli* defective for GreA and GreB have 50% reduced Q-mediated transcription antitermination (111). However, even with reduced antitermination λ can still form plaques on such *greA–greB* mutant strains.

NusA protein stimulates λ Q antitermination at downstream terminators and has an effect very early in stabilizing the binding of Q to the paused complex (185). Results in vitro suggest that NusA displaces the σ subunit from RNA Pol in the transition from initiation to elongation (61, 62, 71). Like NusA, Q binds and displaces σ from the paused complex as the Q-modified termination resistant elongation complex is established (188). Perhaps NusA functions catalytically in helping Q to displace σ from the complex. Still, the mechanism by which NusA stimulates Q is not known. Also, NusA may remain with the Q-RNA Pol elongation complex in a tighter than usual association with RNA Pol.

The combined action of the unique *rpoC10* mutation, altering the β′ subunit of RNA polymerase, with certain mutations in *nusA* has been shown to prevent λ growth and specifically Q antitermination. In this case, Q overproduction reverses the λ defect (85). Q levels are important here, perhaps because there is a competition between Q and NusA for polymerase in the paused complex or because this combination of the defects in β′ and NusA makes displacement of σ more difficult.

Once Q binds to the paused complex it rapidly moves polymerase into the elongation complex and displaces σ (188).

The elegant physical studies of Marr et al. (110) demonstrate the Q-dependent repositioning and displacement of σ-specific binding domains on the DNA at the pause site. This alteration of σ binding by Q presumably weakens σ's tenuous hold on the nontemplate strand, freeing the paused complex to continue elongation. The bound Q protein has replaced σ and has, in some way, still not understood, modified the elongation complex for antitermination. In the new complex, Q is presumed to interact with the core polymerase. Genetic studies suggest that regions in β and β′ may affect this binding as mutations in *rpoB* and *rpoC* affect antitermination action of Q (5, 85).

Once the elongating RNA Pol is modified by Q, its properties change in two ways: it transcribes faster, pausing less on the DNA, and passes through transcription terminators (186, 188). This raises the question as to whether increased transcription speed could cause antitermination. Experiments in which the speed of polymerase is reduced dramatically by limiting substrate nucleotides demonstrate that Q antitermination remains functional. In fact, Q can reduce dissociation of an elongation complex that is stopped at a terminator and would normally be released in the absence of Q. Thus, Q modifies the transcription complex by making it more stable and resistant to dissociation. This complex with increased stability may be able to transcribe DNA more rapidly (188).

In a study looking at both transcription termination and Q antitermination, Yarnell and Roberts (188) demonstrated that a simple antisense oligonucleotide could cause transcription termination of paused polymerase complexes. The antisense segment had to be appropriately located upstream of the pause in the region normally occupied by the stem structure of an intrinsic terminator. When added to static elongation complexes stopped at the pause site, these short antisense DNAs released the RNA and polymerase from the DNA complex. If Q was part of the polymerase complex, Q prevented release just as it can prevent normal termination. Because the antisense oligonucleotide pairing, like the stem-loop structure, is believed to simply dissociate the RNA/DNA hybrid of the paused complex, these authors proposed two general ways Q could work. It could be bound to polymerase upstream of the RNA/DNA hybrid in such a way as to prevent base pairing of the stem structure, or alternatively Q could bind and stabilize the RNA/DNA hybrid itself. Direct stabilization of the RNA/DNA hybrid by Q could also prevent Rho-mediated termination events. Although we do not know how Rho interacts with RNA Pol to cause termination, it must ultimately destabilize the RNA/DNA hybrid (10). We have previously suggested that NusA blocks Rho action under certain conditions. Based on this idea, we further suggest that NusA may serve to facilitate readthrough of Rho termination by Q-modified RNA Pol. According to this idea, Q stabilizes NusA interaction with RNA Pol. Thus, based on our suggestion that the NusA11 protein successfully competes with Rho for binding of RNA Pol (see the section on N action, p. 91), we raise the possibility that Q stabilization of the NusA-RNA Pol interaction facilitates NusA competition with Rho.

Coda

Transcription antitermination can serve two different roles useful to the phage. First, it allows the phage to acquire new genes even if the genes have associated transcription terminators. If an acquired gene locates within an operon, the antitermination mechanism enables transcription to reach downstream genes even if the acquired gene has an associated downstream terminator. For example, in phages carrying genes encoding Shiga toxin, even though the *stx* genes have an associated promoter, p_{stx}, transcription of genes downstream of the *stx* genes only occurs from the $p_{R'}$ promoter, which is upstream of p_{stx} (119, 170). Second, the antitermination system can serve as a regulator of gene expression. The regulation can be temporal and/or physiological.

In considering transcription antitermination in the lambdoid phages, it is curious that these phages employ three different strategies to achieve antitermination. The PUT system appears to rely solely on an RNA structure. The N system relies on RNA sequences and structure as well as both phage- and host-encoded proteins. The Q system relies on a DNA binding site within the promoter, synthesis of a short transcript from that promoter, and a phage and a host protein. What advantages do these different strategies provide to the phage?

Antitermination systems that rely on interactions of sequences in the RNA are always found in the early operons, while those that rely on interactions at sequences in the DNA are found in late operons. We suspect that this arrangement of antitermination sequences is not fortuitous, but reflects some selective advantage to the phage. Regulation of gene expression of early functions is most important in temperate phage development because it is during the expression of these genes that the decision between lysogeny and lysis is made. We have presented studies showing how the N-NUT system provides regulation of gene expression at the levels of transcription and translation. In a subset of phages (those with N-NUT systems like that of H-19B), formation of an RNA structure per se within the NUT region, called the "reducer," appears to contribute to the effectiveness of the NUT site at directing antitermination. When the sequences encoding the reducer structure are removed, the effectiveness of the NUT site in directing N-mediated antitermination is only modestly reduced. In λ, formation of a stem-loop structure per se in the N leader regulates the effectiveness of N-mediated antitermination in another way, by controlling the level of N expression. When present it serves to inhibit expression of N and its removal by RNase III allows higher levels of N to be expressed. Thus, modulation of the action of RNase III provides a mechanism to allow input of the physiological state of the bacterium into the regulation of N expression. The requirement for the action of a number of host proteins also provides additional potential targets for regulatory action.

The situation is different for late gene expression. Once the decision for lysis has been made, monitoring of late gene expression is unnecessary and expression can essentially be constitutive. The level of Q, which is determined by activities regulated by actions in the early operons, determines expression of late genes. Thus, it is likely that proper phage development requires less regulation of late gene expression. Although the λ Q activity is enhanced by NusA in vitro, the activity of the Q of phage 82 is only modestly enhanced by NusA (137, 186). Moreover, none of the other Nus factors appear to contribute to Q action (7). Although NusA may allow some regulation of Q action at *qut*, there is obviously not the panoply of possible regulatory inputs that are available through the NUT RNA site.

Results of studies with phage HK022 appear to contradict the idea that the selective advantage of an RNA antiterminator is that it provides more opportunity for more regulatory inputs. The PUT RNA antiterminator located in the early operons of HK022 appears to function independently of either phage or host auxiliary proteins, unlike the N-Nus modulated NUT RNAs. If the PUT stem-loop and RNA Pol interact in the absence of other inputs and they alone are sufficient to direct processive transcription through downstream terminators, it would be difficult to argue that the antitermination mechanism in the early operon provides an important regulatory capability. PUT RNA is a component of every initiated transcript and therefore it would appear axiomatic that every RNA Pol becomes antitermination-proficient as soon as the *put* sequence is transcribed. In vivo genetic evidence suggests that only the RNA structure is essential for PUT-mediated antitermination. However, two observations showing that PUT-mediated antitermination can be more effective in vivo than in vitro suggest that, as yet, unidentified regulatory controls are exerted on the structure and/or function of PUT in vivo. Firstly, the antitermination activity observed with a DNA template having a *put* site is not as efficient in a pure transcription system with RNA Pol as it is with such a template in vivo. Secondly, some *rpoC* mutations cause a greater decrease in PUT effectiveness in vitro than in vivo (92, 152). Experiments with λ provide direct evidence that N plays a regulatory role in phage development (181). λ forms turbid plaques on *E. coli*, and this indicates that some of the progeny phages formed in the localized infection go down the lysogenic route. However, the same λ forms clear plaques if the *E. coli* lawn constitutively expresses N. This indicates that very few of the progeny phages go down the lysogenic route if N is present from the beginning of the infection. Hence, the timing and perhaps the level of expression of N is an important factor in determining how the phage develops. We have already discussed how expression of N can be regulated; establishing how PUT is regulated or whether it is regulated awaits further work.

While many of the details of the action of λ antiterminators have been elucidated, precisely how they enable RNA Pol to transcend terminators awaits further work. We hope that this chapter has provided new ways to think about this process that may serve to stimulate further experimentation.

Acknowledgments

The authors thank Carolyn McGill, Judah Rosner, Joshua Filter, Robert Weisberg, and Jeffry Roberts for careful reading of parts of or the entire manuscript. Helen Wilson is thanked for help with the illustrations. The work at the University of Michigan was, in part, supported by Public Health Research Grant A111459-10.

References

1. Adhya, S., and M. Gottesman. 1978. Control of transcription termination. Annu. Rev. Biochem. 47:967–996.
2. Adhya, S., M. Gottesman, and B. De Crombrugghe. 1974. Release of polarity in *Escherichia coli* by gene N of phage lambda: termination and antitermination of transcription. Proc. Natl. Acad. Sci. USA 71:2534–2538.
3. Altieri, A. S., M. J. Mazzulla, D. A. Horita, R. H. Coats, P. T. Wingfield, A. Das, D. L. Court, and R. A. Byrd. 2000. The structure of the transcriptional antiterminator NusB from *Escherichia coli*. Nat. Struct. Biol. 7:470–474.
4. Antao, V. P., S. Y. Lai, and I. Tinoco Jr. 1991. A thermodynamic study of unusually stable RNA and DNA hairpins. Nucleic Acids Res. 19:5901–5905.
5. Atkinson, B. L., and M. E. Gottesman. 1992. The *Escherichia coli rpoB60* mutation blocks antitermination by coliphage HK022 Q-function. J. Mol. Biol. 227:29–37.
6. Banik-Maiti, S., R. A. King, and R. A. Weisberg. 1997. The antiterminator RNA of phage HK022. J. Mol. Biol. 272:677–687.
7. Barik, S., and A. Das. 1990. An analysis of the role of host factors in transcription antitermination in vitro by the Q protein of coliphage lambda. Mol. Gen. Genet. 222:152–156.
8. Barik, S., B. Ghosh, W. Whalen, D. Lazinski, and A. Das. 1987. An antitermination protein engages the elongating transcription apparatus at a promoter-proximal recognition site. Cell 50:885–899.
9. Borukhov, S., V. Sagitov, and A. Goldfarb. 1993. Transcript cleavage factors from *E. coli*. Cell 72:459–466.
10. Brennan, C. A., A. J. Dombroski, and T. Platt. 1987. Transcription termination factor rho is an RNA-DNA helicase. Cell 48:945–952.
11. Burns, C. M., L. V. Richardson, and J. P. Richardson. 1998. Combinatorial effects of NusA and NusG on transcription elongation and Rho-dependent termination in *Escherichia coli*. J. Mol. Biol. 278:307–316.

12. Burova, E., and M. E. Gottesman. 1995. NusG overexpression inhibits Rho-dependent termination in *Escherichia coli*. Mol. Microbiol. 17:633–641.
13. Bycroft, M., T. J. P. Hubbard, M. Proctor, S. M. V. Freund, and A. G. Murzin. 1997. The solution structure of the S1 RNA binding domain: a member of an ancient nucleic acid-binding fold. Cell 88:235–242.
14. Cai, Z., A. Gorin, R. Frederick, X. Ye, W. Hu, A. Majumdar, A. Kettani, and D. J. Patel. 1998. Solution structure of P22 transcriptional antitermination N peptide-boxB RNA complex. Nat. Struct. Biol. 5:203–212.
15. Campbell, A. M. 1961. Sensitive mutants of bacteriophage lambda. Virology 14:22–32.
16. Chan, C. L., and R. Landick. 1989. The *Salmonella typhimurium* his operon leader region contains an RNA hairpin-dependent transcription pause site. Mechanistic implications of the effect on pausing of altered RNA hairpins. J. Biol. Chem. 264:20796–20804.
17. Chan, C. L., and R. Landick. 1993. Dissection of the his leader pause site by base substitution reveals a multipartite signal that includes a pause RNA hairpin. J. Mol. Biol. 233:25–42.
18. Chattopadhyay, S., J. Garcia-Mena, J. DeVito, K. Wolska, and A. Das. 1995. Bipartite function of a small RNA hairpin in transcription antitermination in bacteriophage lambda. Proc. Natl. Acad. Sci. USA 92:4061–4065.
19. Chattopadhyay, S., S. C. Hung, A. C. Stuart, A. G. Palmer 3rd, J. Garcia-Mena, A. Das, and M. E. Gottesman. 1995. Interaction between the phage HK022 Nun protein and the nut RNA of phage lambda. Proc. Natl. Acad. Sci. USA 92:12131–12135.
20. Chou, S. H., L. Zhu, and B. R. Reid. 1997. Sheared purine × purine pairing in biology. J. Mol. Biol. 267:1055–1067.
21. Cilley, C. D., and J. R. Williamson. 1997. Analysis of bacteriophage N protein and peptide binding to boxB RNA using polyacrylamide gel coelectrophoresis (PACE). RNA 3:57–67.
22. Clerget, M., D. J. Jin, and R. A. Weisberg. 1995. A zinc-binding region in the β′ subunit of RNA polymerase is involved in antitermination of early transcription of phage HK022. J. Mol. Biol. 248:768–780.
23. Conaway, J. W., A. Shilatifard, A. Dvir, and R. C. Conaway. 2000. Control of elongation by RNA polymerase II. Trends Biochem. Sci. 25:375–380.
24. Court, D. 1993. RNA processing and degradation by RNase III, pp. 71–116. *In* G. Brawerman, and J. Belasco (eds.) Control of mRNA Stability. Academic Press, New York.
25. Court, D. L., T. A. Patterson, N. Baker, N. Costantino, C. Mao, and D. I. Friedman. 1995. Structural and functional analysis of the transcription–translation proteins NusB and NusE. J. Bacteriol. 177:2589–2591.
26. Craven, M. G., and D. I. Friedman. 1991. Analysis of the *Escherichia coli nusA10* (Cs) allele: relating nucleotide changes to phenotypes. J. Bacteriol. 173:1485–1491.
27. Craven, M. G., A. E. Granston, A. T. Schauer, C. Zheng, T. A. Gray, and D. I. Friedman. 1994. *Escherichia coli–Salmonella typhimurium* hybrid *nusA* genes: identification of a short motif required for action of the λ N transcription antitermination protein. J. Bacteriol. 176:1394–1404.
28. Dambly, C., M. Couturier, and R. Thomas. 1968. Control of development in temperate bacteriophages. II. Control of lysozyme synthesis. J. Mol. Biol. 32:67–81.
29. Das, A. 1992. How the phage lambda N gene product suppresses transcription termination: communication of RNA polymerase with regulatory proteins mediated by signals in nascent RNA. J. Bacteriol. 174:6711–6716.
30. Das, A. 1993. Control of transcription termination by RNA-binding proteins. Annu. Rev. Biochem. 63:893–930.
31. Das, A., S. Barik, B. Ghosh, and W. Whalen. 1996. Immunoprinting: a technique used to study dynamic protein–nucleic acid interactions within transcription elongation complex. Methods Enzymol. 274:363–374.
32. Das, A., M. E. Gottesman, J. Wardwell, P. Trisler, and S. Gottesman. 1983. A mutation in the *Escherichia coli rho* gene that inhibits the N protein activity of phage lambda. Proc. Natl. Acad. Sci. USA 80:5530–5534.
33. Das, A., M. Pal, J. G. Mena, W. Whalen, K. Wolska, R. Crossley, W. Rees, P. H. von Hippel, N. Costantino, D. Court, M. Mazzulla, A. S. Altieri, R. A. Byrd, S. Chattopadhyay, J. DeVito, and B. Ghosh. 1996. Components of multiprotein–RNA complex that controls transcription elongation in *Escherichia coli* phage lambda. Methods Enzymol. 274:374–402.
34. Das, A., and K. Wolska. 1984. Transcription antitermination in vitro by lambda *N* gene product: requirement for a phage nut site and the products of host *nusA*, *nusB*, and *nusE* genes. Cell 38:165–173.
35. Dasgupta, S., L. Fernandez, L. Kameyama, T. Inada, Y. Nakamura, A. Pappas, and D. L. Court. 1998. Genetic uncoupling of the dsRNA-binding and RNA cleavage activities of the *Escherichia coli* endoribonuclease RNase III: the effect of dsRNA binding on gene expression. Mol. Microbiol. 28:629–660.
36. De Crombrugghe, B., M. Mudryj, R. DiLauro, and M. Gottesman. 1979. Specificity of the bacteriophage lambda *N* gene product (pN): nut sequences are necessary and sufficient for antitermination by pN. Cell 18:1145–1151.
37. DeVito, J., and A. Das. 1994. Control of transcription processivity in phage lambda: Nus factors strengthen the termination-resistant state of RNA polymerase induced by N antiterminator. Proc. Natl. Acad. Sci. USA 91:8660–8664.
38. Doelling, J. H., and N. C. Franklin. 1989. Effects of all single base substitutions in the loop of boxB on antitermination of transcription by bacteriophage lambda's N protein. Nucleic Acids Res. 17:5565–5577.
39. Dove, W. F. 1966. Action of the lambda chromosome. I. Control of functions late in bacteriophage development. J. Mol. Biol. 19:187–201.
40. Echols, H., D. Court, and L. Green. 1976. On the nature of cis-acting regulatory proteins and genetic organization in bacteriophage: the example of gene Q of bacteriophage lambda. Genetics. 83:5–10.
41. Faber, C., M. Scharpf, T. Becker, H. Sticht, and P. Rosch. 2001. The structure of the coliphage HK022 Nun protein–lambda–phage boxB RNA complex. Implications for the mechanism of transcription termination. J. Biol. Chem. 276:32064–32070.

42. Farnham, P. J., J. Greenblatt, and T. Platt. 1982. Effects of NusA protein on transcription termination in the tryptophan operon of *Escherichia coli*. Cell 29:945–951.
43. Forbes, D., and I. Herskowitz. 1982. Polarity suppression by the *Q* gene product of bacteriophage lambda. J. Mol. Biol. 160:549–569.
44. Franklin, N. C. 1985. Conservation of genome form but not sequence in the transcription antitermination determinants of bacteriophages lambda, phi 21 and P22. J. Mol. Biol. 181:75–84.
45. Franklin, N. C. 1993. Clustered arginine residues of bacteriophage lambda N protein are essential to antitermination of transcription, but their locale cannot compensate for *boxB* loop defects. J. Mol. Biol. 231:343–360.
46. Franklin, N. C., and G. N. Bennett. 1979. The N protein of bacteriophage lambda, defined by its DNA sequence, is highly basic. Gene 8:107–119.
47. Franklin, N. C., and J. H. Doelling. 1989. Overexpression of N antitermination proteins of bacteriophages lambda, phi21, and P22: loss of N protein specificity. J. Bacteriol. 171:2513–2522.
48. Friedman, D. I. 1988. Regulation of phage gene expression by termination and antitermination of transcription, pp. 263–319. *In* R. Calendar (ed.) The Bacteriophage, vol. 2. Plenum Press, New York.
49. Friedman, D. I., M. Baumann, and L. S. Baron. 1976. Cooperative effects of bacterial mutations affecting lambda N gene expression. I. Isolation and characterization of a *nusB* mutant. Virology 73:119–127.
50. Friedman, D. I., and D. L. Court. 1995. Transcription antitermination: the lambda paradigm updated. Mol. Microbiol. 18:191–200.
51. Friedman, D. I., and D. L. Court. 2001. Bacteriophage lambda: alive and well and still doing its thing. Curr. Opin. Microbiol. 4:201–207.
52. Friedman, D. I., and M. Gottesman. 1983. Lytic mode of lambda development, pp. 21–51. *In* R. W. Hendrix, J. W. Roberts, F.W. Stahl, and R. A.Weisberg (eds.) Lambda II. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
53. Friedman, D. I., and E. R. Olson. 1983. Evidence that a nucleotide sequence, "boxA," is involved in the action of the NusA protein. Cell 34:143–149.
54. Friedman, D. I., E. R. Olson, C. Georgopoulos, K. Tilly, I. Herskowitz, and F. Banuett. 1984. Interactions of bacteriophage and host macromolecules in the growth of bacteriophage lambda. Microbiol. Rev. 48:299–325.
55. Friedman, D. I., E. R. Olson, L. L. Johnson, D. Alessi, and M. G. Craven. 1990. Transcription-dependent competition for a host factor: the function and optimal sequence of the phage lambda *boxA* transcription antitermination signal. Genes Dev. 4:2210–2222.
56. Friedman, D. I., A. T. Schauer, M. R. Baumann, L. S. Baron, and S. L. Adhya. 1981. Evidence that ribosomal protein S10 participates in control of transcription termination. Proc. Natl. Acad. Sci. USA 78:1115–1118.
57. Ghosh, B., and A. Das. 1984. nusB: a protein factor necessary for transcription antitermination in vitro by phage lambda N gene product. Proc. Natl. Acad. Sci. USA 81:6305–6309.
58. Ghosh, B., E. Grzadzielska, P. Bhattacharya, E. Peralta, J. DeVito, and A. Das. 1991. Specificity of antitermination mechanisms. Suppression of the terminator cluster T1-T2 of *Escherichia coli* ribosomal RNA operon, *rrnB*, by phage lambda antiterminators. J. Mol. Biol. 222:59–66.
59. Ghysen, A., and M. Pironio. 1972. Relationship between the N function of bacteriophage lambda and host RNA polymerase. J. Mol. Biol. 65:259–272.
60. Gibson, T. J., J. D. Thompson, and J. Heringa. 1993. The KH domain occurs in a diverse set of RNA-binding proteins that include the antiterminator NusA and is probably involved in binding to nucleic acid. FEBS Lett. 324:361–366.
61. Gill, S. C., S. E.Weitzel, and P. H. von Hippel. 1991. *Escherichia coli* sigma 70 and NusA proteins. I. Binding interactions with core RNA polymerase in solution and within the transcription complex. J. Mol. Biol. 220:307–324.
62. Gill, S. C., T. D. Yager, and P. H. von Hippel. 1991. *Escherichia coli* sigma 70 and NusA proteins. II. Physical properties and self-association states. J. Mol. Biol. 220:325–333.
63. Goda, Y., and J. Greenblatt. 1985. Efficient modification of *E. coli* RNA polymerase in vitro by the N gene transcription antitermination protein of bacteriophage lambda. Nucleic Acids Res. 13:2569–2582.
64. Gopal, B., L. F. Haire, R. A. Cox, M. J. Colston, S. Major, J. A. Brannigan, S. J. Smerdon, and G. Dodson. 2000. The crystal structure of NusB from *Mycobacterium tuberculosis*. Nat. Struct. Biol. 7:475–478.
65. Gopal, B., L. F. Haire, S. J. Gamblin, E. J. Dodson, A. N. Lane, K. G. Papavinasasundaram, M. J. Colston, and G. Dodson. 2001. Crystal structure of the transcription elongation/anti-termination factor NusA from *Mycobacterium tuberculosis* at 1.7 Å resolution. J. Mol. Biol. 314:1087–1095.
66. Gopal, B., K. G. Papavinasasundaram, G. Dodson, M. J. Colston, S. A. Major, and A. N. Lane. 2001. Spectroscopic and thermodynamic characterization of the transcription antitermination factor NusE and its interaction with NusB from *Mycobacterium tuberculosis*. Biochemistry 40:920–928.
67. Gottesman, M. E., and R. A. Weisberg. 1995. Termination and antitermination of transcription in temperate bacteriophage. Semin. Virol. 6:35–42.
68. Gottesman, S., M. Gottesman, J. E. Shaw, and M. L. Pearson. 1981. Protein degradation in *E. coli*: the lon mutation and bacteriophage lambda N and cII protein stability. Cell 24:225–233.
69. Grayhack, E. J., X. J. Yang, L. F. Lau, and J. W. Roberts. 1985. Phage lambda gene Q antiterminator recognizes RNA polymerase near the promoter and accelerates it through a pause site. Cell 42:259–269.
70. Greenblatt, J. 1992. Protein–protein interactions as critical determinants of regulated initiation and termination of transcription, pp. 203–226. *In* S. L. McKnight, and K. R. Yamamoto (eds.) Transcriptional Regulation. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
71. Greenblatt, J., and J. Li. 1981. Interaction of the sigma factor and the *nusA* gene protein of *E. coli* with RNA polymerase in the initiation–termination cycle of transcription. Cell 24:421–428.

72. Greenblatt, J., and J. Li. 1982. Properties of the N gene transcription antitermination protein of bacteriophage lambda. J. Biol. Chem. 257:362–365.
73. Greenblatt, J., T. F. Mah, P. Legault, J. Mogridge, J. Li, and L. E. Kay. 1998. Structure and mechanism in transcriptional antitermination by the bacteriophage lambda N protein. Cold Spring Harb. Symp. Quant. Biol. LXIII:327–336.
74. Greenblatt, J., J. R. Nodwell, and S. W. Mason. 1993. Transcriptional antitermination. Nature 364:401–406.
75. Gross, C. A., C. Chan, A. Dombroski, T. Gruber, M. Sharp, J. Tupy, and B. Young. 1998. The functional and regulatory roles of sigma factors in transcription. Cold Spring Harb. Symp. Quant. Biol. LXIII:141–155.
76. Gusarov, I., and E. Nudler. 2001. Control of intrinsic transcription termination by N and NusA: the basic mechanisms. Cell 107:437–449.
77. Henthorn, K. S., and D. I. Friedman. 1996. Identification of functional regions of the Nun transcription termination protein of phage HK022 and the N antitermination protein of phage gamma using hybrid *nun-N* genes. J. Mol. Biol. 257:9–20.
78. Herskowitz, I., and D. Hagen. 1980. The lysis–lysogeny decision of phage lambda: explicit programming and responsiveness. Annu. Rev. Genet. 14:399–445.
79. Heus, H. A., and A. Pardi. 1991. Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. Science 253:191–194.
80. Hilliker, S., and D. Botstein. 1975. An early regulatory gene of *Salmonella* phage P22 analogous to gene N of coliphage lambda. Virology 68:510–524.
81. Horwitz, R. J., J. Li, and J. Greenblatt. 1987. An elongation control particle containing the N gene transcriptional antitermination protein of bacteriophage lambda. Cell 51:631–641.
82. Huang, A., J. Friesen, and J. L. Brunton. 1987. Characterization of a bacteriophage that carries the genes for production of Shiga-like toxin I in *Escherichia coli*. J. Bacteriol. 169:4308–4312.
83. Hung, S. C., and M. E. Gottesman. 1995. Phage HK022 Nun protein arrests transcription on phage λ DNA in vitro and competes with the phage λ N antitermination protein. J. Mol. Biol. 247:428–442.
84. Hung, S. C., and M. E. Gottesman. 1997. The Nun protein of bacteriophage HK022 inhibits translocation of *Escherichia coli* RNA polymerase without abolishing its catalytic activities. Genes Dev. 11:2670–2678.
85. Ito, K., and Y. Nakamura. 1993. Pleiotropic effects of the *rpoC10* mutation affecting the RNA polymerase β′ subunit of *Escherichia coli* on factor-dependent transcription termination and antitermination. Mol. Microbiol. 9:285–293.
86. Jin, D. J., M. Cashel, D. I. Friedman, Y. Nakamura, W. A. Walter, and C. A. Gross. 1988. Effects of rifampicin resistant *rpoB* mutations on antitermination and interaction with nusA in *Escherichia coli*. J. Mol. Biol. 204:247–261.
87. Joyner, A., L. N. Isaacs, H. Echols, and W. S. Sly. 1966. DNA replication and messenger RNA production after induction of wild-type lambda bacteriophage and lambda mutants. J. Mol. Biol. 19:174–186.
88. Juhala, R. J., M. E. Ford, R. L. Duda, A. Youlton, G. F. Hatfull, and R. W. Hendrix. 2000. Genomic sequences of bacteriophages HK97 and HK022: pervasive mosaicism in the lambdoid phages. J. Mol. Biol. 299:27–51.
89. Kainz, M., and J. Roberts. 1992. Structure of transcription elongation complexes in vivo. Science 255:838–841.
90. Kameyama, L., L. Fernandez, D. L. Court, and G. Guarneros. 1991. RNaseIII activation of bacteriophage lambda N synthesis. Mol. Microbiol. 5:2953–2963.
91. Keppel, F., C. P. Georgopoulos, and H. Eisen. 1974. Host interference with expression of the lambda N gene product. Biochimie 56:1505–1509.
92. King, R. A., S. Banik-Maiti, D. J. Jin, and R. A. Weisberg. 1996. Transcripts that increase the processivity and elongation rate of RNA polymerase. Cell 87:893–903.
93. King, R. A., P. L. Madsen, and R. A. Weisberg. 2000. Constitutive expression of a transcription termination factor by a repressed prophage: promoters for transcribing the phage HK022 *nun* gene. J. Bacteriol. 182:456–462.
94. Ko, D. C., M. T. Marr, J. Guo, and J. W. Roberts. 1998. A surface of *Escherichia coli* sigma 70 required for promoter function and antitermination by phage lambda Q protein. Genes Dev. 12:3276–3285.
95. Komissarova, N., and M. Kashlev. 1997. RNA polymerase switches between inactivated and activated states by translocating back and forth along the DNA and the RNA. J. Biol. Chem. 272:15329–15338.
96. Korzheva, N., A. Mustaev, E. Nudler, V. Nikiforov, and A. Goldfarb. 1998. Mechanistic model of the elongation complex of *Escherichia coli* RNA polymerase. Cold Spring Harb. Symp. Quant. Biol. LXIII:337–435.
97. Landick, R. 1997. RNA polymerase slides home: pause and termination site recognition. Cell 88:741–744.
98. Lau, L. F., and J. W. Roberts. 1985. Rho-dependent transcription termination at lambda tR1 requires upstream sequences. J. Biol. Chem. 260:574–584.
99. Lazinski, D., E. Grzadzielska, and A. Das. 1989. Sequence-specific recognition of RNA hairpins by bacteriophage antiterminators requires a conserved arginine-rich motif. Cell 59:207–218.
100. Lecocq, J., and C. Dambly. 1976. A bacterial RNA polymerase mutant that renders lambda growth independent of the N and *cro* functions at 42 degrees C. Mol. Gen. Genet. 145:53–64.
101. Legault, P., J. Li, J. Mogridge, L. E. Kay, and J. Greenblatt. 1998. NMR structure of the bacteriophage lambda N peptide/boxB RNA complex: recognition of a GNRA fold by an arginine-rich motif. Cell 93:289–299.
102. Lew, K., and S. Casjens. 1975. Identification of early proteins coded by bacteriophage P22. Virology 68:525–533.
103. Li, J., R. Horwitz, S. McCracken, and J. Greenblatt. 1992. NusG, a new *Escherichia coli* elongation factor involved in transcriptional antitermination by the N protein of phage lambda. J. Biol. Chem. 267:6012–6019.
104. Li, J., S. W. Mason, and J. Greenblatt. 1993. Elongation factor NusG interacts with termination factor rho to

regulate termination and antitermination of transcription. Genes Dev. 7:161–172.

105. Liu, K., and M. M. Hanna. 1995. NusA interferes with interactions between the nascent RNA and the C-terminal domain of the alpha subunit of RNA polymerase in *Escherichia coli* transcription complexes. Proc. Natl. Acad. Sci. USA 92:5012–5016.

106. Liu, K., Y. Zhang, K. Severinov, A. Das, and M. M. Hannah. 1996. Role of *Escherichia coli* RNA polymerase alpha subunit in modulation of pausing, termination and antitermination by the transcription elongation factor NusA. EMBO J. 15:150–161.

107. Lozeron, H. A., J. E. Dahlberg, and W. Szybalski. 1976. Processing of the major leftward mRNA of coliphage lambda. Virology 71:262–277.

108. Luttgen, H., R. Robelek, R. Muhlberger, T. Diercks, S. C. Schuster, P. Kohler, H. Kessler, A. Bacher, and G. Richter. 2002. Transcriptional regulation by antitermination. Interaction of RNA with NusB Protein and NusB/NusE protein complex of *Escherichia coli*. J. Mol. Biol. 316:875–885.

109. Mah, T. F., J. Li, A. R. Davidson, and J. Greenblatt. 1999. Functional importance of regions in *Escherichia coli* elongation factor NusA that interact with RNA polymerase, the bacteriophage lambda N protein and RNA. Mol. Microbiol. 34:523–537.

110. Marr, M. T., S. A. Datwyler, C. F. Meares, and J. W. Roberts. 2001. Restructuring of an RNA polymerase holoenzyme elongation complex by lambdoid phage Q proteins. Proc. Natl. Acad. Sci. USA 98:8972–8978.

111. Marr, M. T., and J. W. Roberts. 2000. Function of transcription cleavage factors GreA and GreB at a regulatory pause site. Mol. Cell 6:1275–1285.

112. Mason, S. W., J. Li, and J. Greenblatt. 1992. Direct interaction between two *Escherichia coli* transcription antitermination factors, NusB and ribosomal protein S10. J. Mol. Biol. 223:55–66.

113. Mason, S. W., J. Li, and J. Greenblatt. 1992. Host factor requirements for processive antitermination of transcription and suppression of pausing by the N protein of bacteriophage lambda. J. Biol. Chem. 267:19418–19426.

114. Mogridge, J., P. Legault, J. Li, M. D. Van Oene, L. E. Kay, and J. Greenblatt. 1998. Independent ligand-induced folding of the RNA–binding domain and two functionally distinct antitermination regions in the phage lambda N protein. Mol. Cell 1:265–275.

115. Mogridge, J., T.-F. Mah, and J. Greenblatt. 1995. A protein–RNA interaction network facilitates the template-independent cooperative assembly on RNA polymerase of a stable antitermination complex containing the λ N protein. Genes Dev. 9:2831–2844.

116. Mogridge, J., T. F. Mah, and J. Greenblatt. 1998. Involvement of boxA nucleotides in the formation of a stable ribonucleoprotein complex containing the bacteriophage lambda N protein. J. Biol. Chem. 273:4143–4148.

117. Mooney, R. A., I. Artsimovitch, and R. Landick. 1998. Information processing by RNA polymerase: recognition of regulatory signals during RNA chain elongation. J. Bacteriol. 180:3265–3275.

118. Nakamura, Y., S. Mizusawa, A. Tsugawa, and M. Imai. 1986. Conditionally lethal *nusAts* mutation of *Escherichia coli* reduces transcription termination but does not affect antitermination of bacteriophage lambda. Mol. Gen. Genet. 204:24–28.

119. Neely, M. N., and D. I. Friedman. 1998. Functional and genetic analysis of regulatory regions of coliphage H-19B: location of Shiga-like toxin and lysis genes suggest a role for phage functions in toxin release. Mol. Microbiol. 28:1255–1267.

120. Neely, M. N., and D. I. Friedman. 2000. N-mediated transcription antitermination in lambdoid phage H-19B is characterized by alternative NUT RNA structures and a reduced requirement for host factors. Mol. Microbiol. 38:1074–1085.

121. Nehrke, K. W., and T. Platt. 1994. A quaternary transcription termination complex. Reciprocal stabilization by Rho factor and NusG protein. J. Mol. Biol. 243:830–839.

122. Nodwell, J. R., and J. Greenblatt. 1993. Recognition of boxA antiterminator RNA by the *E. coli* antitermination factors NusB and ribosomal protein S10. Cell 72:261–268.

123. Nudler, E. 1999. Transcription elongation: structural basis and mechanisms. J. Mol. Biol. 288:1–12.

124. Oberto, J., M. Clerget, M. Ditto, K. Cam, and R. A. Weisberg. 1993. Antitermination of early transcription in phage HK022. Absence of a phage-encoded antitermination factor. J. Mol. Biol. 229:368–381.

125. Obuchowski, M., A. Wegrzyn, A. Szalewska-Palasz, M. S. Thomas, and G. Wegrzyn. 1997. An RNA polymerase alpha subunit mutant impairs N-dependent transcriptional antitermination in *Escherichia coli*. Mol. Microbiol. 23:211–222.

126. Olson, E. R., E. L. Flamm, and D. I. Friedman. 1982. Analysis of *nutR*: a region of phage lambda required for antitermination of transcription. Cell 31:61–70.

127. Olson, E. R., C. S. Tomich, and D. I. Friedman. 1984. The *nusA* recognition site. Alteration in its sequence or position relative to upstream translation interferes with the action of the N antitermination function of phage lambda. J. Mol. Biol. 180:1053–1063.

128. Pasman, Z., and P. H. von Hippel. 2000. Regulation of rho-dependent transcription termination by NusG is specific to the *Escherichia coli* elongation complex. Biochemistry 39:5573–5585.

129. Patterson, T. A., Z. Zhang, T. Baker, L. L. Johnson, D. I. Friedman, and D. L. Court. 1994. Bacteriophage lambda N-dependent transcription antitermination. Competition for an RNA site may regulate antitermination. J. Mol. Biol. 236:217–228.

130. Plunkett, G. 3rd, D. J. Rose, T. J. Durfee, and F. R. Blattner. 1999. Sequence of Shiga toxin 2 phage 933W from *Escherichia coli* O157:H7: Shiga toxin as a phage late-gene product. J. Bacteriol. 181:1767–1778.

131. Ptashne, M. 1992. A Genetic Switch, 2nd edn. Cell Press, Blackwell Scientific, Cambridge, Mass.

132. Rees, W. A., S. E. Weitzel, T. D. Yager, A. Das, and P. H. von Hippel. 1996. Bacteriophage lambda N protein alone can induce transcription antitermination in vitro. Proc. Natl. Acad Sci. USA 93:342–346.

133. Richardson, J. P., and J. Greenblatt. 1996. Control of RNA chain elongation and termination, pp. 822–848. *In* F. C. Neidhardt (ed.) *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology. American Society for Microbiology, Washington, D.C.
134. Ring, B. Z., and J. W. Roberts. 1994. Function of a nontranscribed DNA strand site in transcription elongation. Cell 78:317–324.
135. Ring, B. Z., W. S. Yarnell, and J. W. Roberts. 1996. Function of *E. coli* RNA polymerase sigma factor sigma 70 in promoter-proximal pausing. Cell 86:485–493.
136. Robert, J., S. B. Sloan, R. A. Weisberg, M. E. Gottesman, R. Robledo, and D. Harbrecht. 1987. The remarkable specificity of a new transcription termination factor suggests that the mechanisms of termination and antitermination are similar. Cell 51:483–492.
137. Roberts, J. 1992. Antitermination and the control of transcription elongation, pp. 389–406. *In* S. L. McKnight and K. R. Yamamoto (eds.) Transcription Regulation. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
138. Roberts, J. W. 1969. Termination factor for RNA synthesis. Nature 224:1168–1174.
139. Roberts, J. W. 1975. Transcription termination and late control in phage lambda. Proc. Natl. Acad. Sci. USA. 72:3300–3304.
140. Roberts, J. W. 1976. Transcription termination and its control in *E. coli*, pp. 247–271. *In* R. Losick and M. Chamberlin (eds.), RNA Polymerase. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
141. Roberts, J. W. 1993. RNA and protein elements of *E. coli* and lambda transcription antitermination complexes. Cell 72:653–655.
142. Roberts, J. W., C. W. Roberts, S. Hilliker, and D. Botstein. 1976. Transcription termination and Regulation in bacteriophages P22 and lambda, pp. 707–718. *In* R. Losick and M. Chamberlin (eds.) RNA Polymerase. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
143. Roberts, J. W., W. Yarnell, E. Bartlett, J. Guo, M. Marr, D. C. Ko, H. Sun, and C. W. Roberts. 1998. Antitermination by bacteriophage lambda Q protein. Cold Spring Harb. Symp. Quant. Biol. LXIII:319–325.
144. Robledo, R., B. L. Atkinson, and M. E. Gottesman. 1991. *Escherichia coli* mutations that block transcription termination by phage HK022 Nun protein. J. Mol. Biol. 220:613–619.
145. Robledo, R., M. E. Gottesman, and R. A. Weisberg. 1990. λ nutR mutations convert HKO22 Nun protein from a transcription termination factor to a suppressor of termination. J. Mol. Biol. 212:635–643.
146. Rosenberg, M., D. Court, H. Shimatake, C. Brady, and D. L. Wulff. 1978. The relationship between function and DNA sequence in an intercistronic regulatory region in phage lambda. Nature 272:414–423.
147. Ruteshouser, E. C., and J. P. Richardson. 1989. Identification and characterization of transcription termination sites in the *Escherichia coli lacZ* gene. J. Mol. Biol. 208:23–43.
148. Salstrom, J. S., and W. Szybalski. 1978. Coliphage lambda nutL$^-$: a unique class of mutants defective in the site of gene N product utilization for antitermination of leftward transcription. J. Mol. Biol. 124:195–221.
149. Scharpf, M., H. Sticht, K. Schweimer, M. Boehm, S. Hoffmann, and P. Rosch. 2000. Antitermination in bacteriophage lambda. The structure of the N36 peptide–boxB RNA complex. Eur. J. Biochem. 267:2397–2408.
150. Schauer, A. T., D. L. Carver, B. Bigelow, L. S. Baron, and D. I. Friedman. 1987. Lambda N antitermination system: functional analysis of phage interactions with the host NusA protein. J. Mol. Biol. 194:679–690.
151. Schauer, A. T., S. C. Cheng, C. Zheng, L. St. Pierre, D. Alessi, D. L. Hidayetoglu, N. Costantino, D. L. Court, and D. I. Friedman. 1996. The alpha subunit of RNA polymerase and transcription antitermination. Mol. Microbiol. 21:839–851.
152. Sen, R., R. A. King, N. Mzhavia, P. L. Madsen, and R. A. Weisberg. 2002. Sequence specific interaction of nascent antitermination RNA with the zinc finger motif of *E. coli* RNA polymerase Mol. Microbiol. 46:215–222.
153. Sen, R., R. A. King, and R. A. Weisberg. 2001. Modification of the properties of elongating RNA polymerase by persistent association with nascent antiterminator RNA. Mol. Cell, 7:993–1001.
154. Sharrock, R. A., R. L. Gourse, and M. Nomura. 1985. Defective antitermination of rRNA transcription and derepression of rRNA and tRNA synthesis in the *nusB5* mutant of *Escherichia coli*. Proc. Natl. Acad. Sci. USA 82:5275–5279.
155. Sidorenkov, I., N. Komissarova, and M. Kashlev. 1998. Crucial role of the RNA:DNA hybrid in the processivity of transcription. Mol. Cell 2:55–64.
156. Sigmund, C. D., and E. A. Morgan. 1988. NusA protein affects transcriptional pausing and termination in vitro by binding to different sites on the transcription complex. Biochemistry 27:5622–5627.
157. Sloan, S. B., and R. A. Weisberg. 1993. Use of a gene encoding a suppressor tRNA as a reporter of transcription: analyzing the action of the Nun protein of bacteriophage HK022. Proc. Natl. Acad. Sci. USA 90:9842–9846.
158. Smith, H. W., P. Green, and Z. Parsell. 1983. Vero cell toxins in *Escherichia coli* and related bacteria: transfer by phage and conjugation and toxic action in laboratory animals, chickens and pigs. J. Gen. Microbiol. 129:3121–3137.
159. Somasekhar, G., and W. Szybalski. 1983. Mapping of the Q-utilization site (*qut*) required for antitermination of late transcription in bacteriophage lambda. Gene 26:291–294.
160. Somasekhar, G., and W. Szybalski. 1987. The functional boundaries of the Q-utilization site required for antitermination of late transcription in bacteriophage lambda. Virology 158:414–426.
161. Squires, C. L., J. Greenblatt, J. Li, and C. Condon. 1993. Ribosomal RNA antitermination in vitro: requirement for Nus factors and one or more unidentified cellular components. Proc. Natl. Acad. Sci. USA 90:970–974.
162. Su, L., J. T. Radek, L. A. Labeots, K. Hallenga, P. Hermanto, H. Chen, S. Nakagawa, M. Zhao, S. Kates, and M. A. Weiss. 1997. An RNA enhancer in a phage transcriptional antitermination complex functions as a structural switch. Genes Dev. 11:2214–2226.

163. Sullivan, S. L., and M. E. Gottesman. 1992. Requirement for *E. coli* NusG protein in factor-dependent transcription termination. Cell 68:989–994.
164. Sullivan, S. L., D. F. Ward, and M. E. Gottesman. 1992. Effect of *Escherichia coli* nusG function on lambda N-mediated transcription antitermination. J. Bacteriol. 174:1339–1344.
165. Tan, R., and A. D. Frankel. 1995. Structural variety of arginine–rich RNA-binding peptides. Proc. Natl. Acad. Sci. USA 92:5282–5286.
166. Taura, T., C. Ueguchi, K. Shiba, and K. Ito. 1992. Insertional disruption of the *nusB* (*ssyB*) gene leads to cold-sensitive growth of *Escherichia coli* and suppression of the *secY24* mutation. Mol. Gen. Genet. 234:429–432.
167. Van Gilst, M. R., W. A. Rees, A. Das, and P. H. von Hippel. 1997. Complexes of N antitermination protein of phage lambda with specific and nonspecific RNA target sites on the nascent transcript. Biochemistry 36:1514–1524.
168. Van Gilst, M. R., and P. H. von Hippel. 1997. Assembly of the N-dependent antitermination complex of phage lambda: NusA and RNA bind independently to different unfolded domains of the N protein. J. Mol. Biol. 274:160–173.
169. Vogel, U., and K. F. Jensen. 1997. NusA is required for ribosomal antitermination and for modulation of the transcription elongation rate of both antiterminated RNA and mRNA. J. Biol. Chem. 272:12265–12271.
170. Wagner, P. L., M. N. Neely, X. Zhang, D. W. Acheson, M. K. Waldor, and D. I. Friedman. 2001. Role for a phage promoter in Shiga toxin 2 expression from a pathogenic *Escherichia coli* strain. J. Bacteriol. 183:2081–2085.
171. Ward, D. F., A. DeLong, and M. E. Gottesman. 1983. *Escherichia coli nusB* mutations that suppress *nusA1* exhibit lambda N specificity. J. Mol. Biol. 168:73–85.
172. Ward, D. F., and M. E. Gottesman. 1981. The nus mutations affect transcription termination in *Escherichia coli*. Nature 292:212–215.
173. Warren, F., and A. Das. 1984. Formation of termination-resistant transcription complex at phage lambda nut locus: effects of altered translation and a ribosomal mutation. Proc. Natl. Acad. Sci. USA 81:3612–3616.
174. Watnick, R. S., and M. E. Gottesman. 1998. *Escherichia coli NusA is required for efficient RNA binding by phage HK022 nun protein. Proc. Natl. Acad Sci. USA 95:1546–1551.*
175. Watnick, R. S., and M. E. Gottesman. 1999. Binding of transcription termination protein nun to nascent RNA and template DNA. Science 286:2337–2339.
176. Watnick, R. S., S. C. Herring, A. G. Palmer 3rd, and M. E. Gottesman. 2000. The carboxyl terminus of phage HK022 Nun includes a novel zinc-binding motif and a tryptophan required for transcription termination. Genes Dev. 14:731–739.
177. Weisberg, R. A., and M. E. Gottesman. 1999. Processive antitermination. J. Bacteriol. 181:359–367.
178. Weisberg, R. A., M. E. Gottesman, R. W. Hendrix, and J. W. Little. 1999. Family values in the age of genomics: comparative analysis of temperate bacteriophage HK022. Annu. Rev. Genet. 33:565–602.
179. Whalen, W., B. Ghosh, and A. Das. 1988. NusA protein is necessary and sufficient in vitro for phage lambda N gene product to suppress a rho-independent terminator placed downstream of nutL. Proc. Natl. Acad. Sci. USA 85:2494–2498.
180. Whalen, W. A., and A. Das. 1990. Action of an RNA site at a distance: role of the nut genetic signal in transcription antitermination by phage-lambda N gene product. New Biol. 2:975–991.
181. Wilson, H. R., L. Kameyama, J. G. Zhou, G. Guarneros, and D. L. Court. 1997. Translational repression by a transcriptional elongation factor. Genes Dev. 11:2204–2213.
182. Wilson, H. R., D. Yu, H. K. Peters 3rd, J. G. Zhou, and D. L. Court. 2002. The global regulator RNase III modulates translation repression by the transcription elongation factor N. EMBO J. 21:4154–4161.
183. Worbs, M., G. P. Bourenkov, H. D. Bartunik, R. Huber, and M. C. Wahl. 2001. An extended RNA binding surface through arrayed S1 and KH domains in transcription factor NusA. Mol. Cell 7:1177–1189.
184. Yager, T. D., and P. H. Von Hippel. 1987. Transcript elongation and termination in *Escherichia coli*, pp. 1241–1275. *In* F. C. Neidhardt, J. L. Ingraham, B. Magasanik, K. B. Low, M. Schaechter, and H. E. Umbarger (eds.), *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology. ASM, Washington, D.C.
185. Yang, X. J., C. M. Hart, E. J. Grayhack, and J. W. Roberts. 1987. Transcription antitermination by phage lambda gene Q protein requires a DNA segment spanning the RNA start site. Genes Dev. 1:217–226.
186. Yang, X. J., and J. W. Roberts. 1989. Gene Q antiterminator proteins of *Escherichia coli* phages 82 and lambda suppress pausing by RNA polymerase at a rho-dependent terminator and at other sites. Proc. Natl. Acad. Sci. USA 86:5301–5305.
187. Yarnell, W. S., and J. W. Roberts. 1992. The phage lambda gene Q transcription antiterminator binds DNA in the late gene promoter as it modifies RNA polymerase. Cell 69:1181–1189.
188. Yarnell, W. S., and J. W. Roberts. 1999. Mechanism of intrinsic transcription termination and antitermination. Science 284:611–615.
189. Zheng, C., and D.I. Friedman. 1994. Reduced Rho-dependent transcription termination permits NusA-independent growth of *Escherichia coli*. Proc. Natl. Acad. Sci. USA 91:7543–7547.
190. Zhou, Y., J. J. Filter, D. L. Court, M. E. Gottesman, and D. I. Friedman. 2002. A Requirement for NusG for transcription antitermination in vivo by the λ N protein. J. Bacteriol. 184:3416–3418.
191. Zhou, Y., T. F. Mah, J. Greenblatt, and D. I. Friedman. 2002. Evidence that the KH RNA-binding domains influence action of the *E. coli* NusA protein. J. Mol. Biol. 318:1175–1188.
192. Zhou, Y., T. F. Mah, Y. T. Yu, J. Mogridge, E. R. Olson, J. Greenblatt, and D. I. Friedman. 2001. Interactions of an Arg-rich region of transcription elongation protein NusA with NUT RNA: implications for the order of assembly of the lambda N antitermination complex in vivo. J. Mol. Biol. 310:33–49.

10

Phage Lysis

RY YOUNG
ING-NANG WANG

Timing Is Everything

In general, bacteriophages must lyse their host cells to liberate the progeny virions. The most common misconception about host lysis is that it is simply the inevitable outcome of phage infection, a kind of sad denouement that follows after the elegant processes defining successive waves of gene expression and the intricate and coordinated pathways leading to virion morphogenesis. For example, in the otherwise captivating description of the molecular processes of phage life cycles in *The Molecular Biology of the Gene* (112), it is explained that some phage proteins are needed early, "while others, such as the viral coat proteins or the lysozyme for bacterial cell lysis, must only appear later," leading to the end of the infective cycle of phage T7. "Finally, T7 lysozyme destroys the bacterial cell wall and allows newly formed phage particles to escape and renew their growth cycle in other bacteria." The impression given is that lysis is caused by the late appearance of lysozyme activity, despite the fact that T7 lysozyme is an *early* gene. Also, the implication is that release and reinfection is needed to "renew" a completed vegetative program, that lysis occurs *after* virion production had been achieved. However, it has been known for decades that if the lysis genes are ablated, the accumulation of intracellular virions can continue unabated for extended periods, far beyond the time of lysis of parental phage (67). In fact, a recent re-analysis of the kinetics of virion accumulation for lysis-defective mutant phages indicated that it is essentially linear for periods at least 4-fold longer than the wild-type vegetative cycle (110). Thus lysis is a *programmed event*, like all the other steps in the infectious cycle. It is the thesis here that the decision of when to terminate the infection and lyse the host is the *only* major decision made in the vegetative cycle; that is, that all other macromolecular processes proceed at rates which maximize the intracellular accumulation of infective virions.

Lysis Systems: At Least Two Strategies

Most bacteria have a murein cell wall that constitutes the principal barrier to host lysis. Compromising the cell wall is thus the fundamental goal for lytic processes. Two general strategies to accomplish lysis have been described (121). Phages with double-stranded nucleic acid genomes use the "holin–endolysin" strategy. In this scheme, the phage elaborates a murein-degrading enzyme, an *endolysin* (also known as lysozyme), specifically dedicated to lysis, and a second, membrane-embedded protein, the *holin*, which serves to activate the endolysin at a precisely defined time. This is the basis of the programmed nature of host lysis. As the principal scheduling component, the holin is subject to elaborate regulatory schemes in real time and to exacting evolutionary pressures. This chapter will focus heavily on the remarkable properties of holins, which can be considered the smallest and simplest biological timers.

In contrast, small single-stranded nucleic acid phages, with very limited genomes, achieve lysis without encoding a muralytic enzyme. For phage biologists, this "lysis-without-lysozyme" has been a long-standing mystery that now, for at least two such phages, has been solved, at least in outline. In the cases of the ssDNA phage ϕX174 and the ssRNA phage Qβ, a single phage protein, for which we have proposed the term "amurin" (14), causes lysis by acting as a specific inhibitor of an enzyme in the multi-step pathway of murein biosynthesis (10, 12, 14). Because this kind of host lysis requires continued cell growth and appears to involve ruptures of the wall at the developing septum, just as in the case of cell wall antibiotics like penicillin, the amurins can be considered "protein antibiotics." Remarkably, the enzymatic steps in the murein synthesis pathway that are inhibited by the ϕX174 and Qβ amurins are different, and there is evidence that other small phages attack other steps (T. G. Bernhardt, B. McIntosh, and R. Young, unpublished data), raising the prospect that in the largely unexplored universe of small phages, there may be an amurin specific for many of the steps in the widely conserved pathway of peptidoglycan synthesis.

Holin–Endolysin Lysis: The Universal Strategy for Complex Phages

Why Did Holins Evolve?

In principle, lysis requires only that the cell wall be at least partly degraded, or that its synthesis be corrupted during growth. Thus, to achieve lysis a phage needs only to produce a muralytic enzyme with a secretory signal sequence or to generate a polypeptide that blocks murein biosynthesis. However, it appears that all phages with sufficient genomic space have evolved a more sophisticated lysis system, involving at least one other protein, the holin, which imposes a strict temporal program on the muralytic enzyme, the endolysin. The universality of the holin program, ensuring that lysis occurs at a programmed time despite the undiminished capacity of the infected cell for continued virion production, must reflect an evolutionary adaptation. Theoretical analysis suggests that, for any specified host and environmental scenario, there is an ideal or optimal lysis time (1, 110; chapter 5). Irrespective of the details of the calculation, the clarifying perspective is simple to state: continued linear accumulation of virions in the infected host must be balanced against the potential for exponential accumulation if the progeny virions are released to pursue new prey. Thus a lysis system in which the programmed time of lysis can be easily adjusted in response to the imposition of different physiological and environmental conditions would contribute strongly to fitness. This adjustability could be reflected in terms of genetic variability of the timing, such that most missense changes in the timing gene would cause profound changes in the lysis time, allowing alleles with the appropriate lysis timing to arise rapidly, or it might take the form of real-time regulation of the lysis event, using environmental signals as inputs to key the programming of lysis.

What characteristics would be appropriate for a system evolved to effect host lysis at a defined time? First, the lysis system should be as saltatory as possible; that is, the lysis system should not affect the productivity of the infected host, in terms of virion assembly, until the programmed time of lysis. Second, the lysis system should be very efficient and rapid once the infective cycle is terminated. That is, the time between the cessation of host macromolecular synthesis and the bursting of the cell should be minimal; there is no profit from dwell time in a corpse. Finally, the timer should be capable of being over-ridden in real time, should circumstances change during an infection. For example, if environmental signals are received indicating a lack of available hosts in the environment, the timer should be blocked from triggering lysis so that at least the continued linear accumulation of virions can be maintained. Conversely, if there is a sudden loss of energy or a superinfection by a heterologous phage, the timer should be triggered to allow immediate lysis ("bail out").

All these characteristics are exhibited by the holin–endolysin system of lysis and are mostly due to the properties of holins alone.

Definitions and Significance

Holins are small, phage-encoded membrane proteins which accumulate in the cytoplasmic membrane of the host. At a precise time programmed into its primary structure, the holin suddenly causes disruption of the membrane, or "hole formation." This terminology was chosen deliberately because the word "hole" has an appropriately vague connotation not connected with a specific membrane structure. Also, the use of this term avoids confusion with porins, which are well-defined proteins in the outer membrane of Gram-negative bacteria. Hole formation leads directly, and usually very rapidly, to destruction of the cell wall by the phage-encoded muralytic enzymes, or endolysins. Although recent evidence suggests that considering endolysins as passive, uninteresting reporter enzymes for holin activity may be too simplistic (see Emerging Perspectives, below), much of this chapter is devoted to documenting the molecular basis for the remarkable capacities of holins, the simplest of biological "timers." Again, this choice of terminology is calculated, because holins are not biological clocks, in the classic sense. Biological clocks run continuously and may be referenced by many developmental processes, whereas holins run only once, act on their own to impose a precise temporal limit on the phage infective cycle, and, insofar as is known, have no other effect on the sequencing of the macromolecular events of virion morphogenesis. This conclusion in no way diminishes the importance or significance of holins in the biosphere. Phage predation of bacteria accounts for half or more of the flux of carbon through the world's total marine biomass, where 50 Gt of C is fixed annually (44, 45, 113). As noted above, the only actively controlled variable in lytic growth is the length of the vegetative cycle; thus, this most ancient temporal regulation is a profoundly important factor for life on Earth. As befits ancient proteins with such a heavy responsibility, holins are incredibly diverse, perhaps the most diverse group of proteins with common function in biology.

It is worth noting that the term "endolysin" is used for phage muralytic enzymes because of its place in a historically important debate over whether the enzyme activities that appear during viral infections are virus-encoded or induced from the host chromosome (36). Moreover, the more familiar name *lysozyme* has been used as the name for the muralytic enzymes of λ, T4, and T7, despite the fact that the three proteins have different enzymatic activities and only the T4 lysozyme has the same enzymatic action as the generic egg white lysozyme.

The λ Paradigm

At the time of the textbook citation above, phage genetics had already revealed that the muralytic enzyme was not the only required lysis protein. In three different systems—λ (49, 57), P22 (35), and T4 (67)—phage mutants had been identified which not only failed to cause lysis but continued to accumulate virions and lysozyme intracellularly far beyond the wild-type lysis time. These mutations defined λ *S*, its close relative P22 *13*, and the unrelated T4 *t*, all of which encode *holins*. The phenotypes highlight a characteristic unique to holin genes: they are the only genes with greatly *increased* production of virions as part of their null phenotypes (111)!

The paradigm holin–endolysin lysis system is that of phage λ. The λ lysis cassette, immediately downstream of the late promoter, has an unusually rich architecture, consisting of three tandem cistrons, *SRRz* (figure 10-1A), which give rise to five protein products. One of the extra products is encoded by a fourth cistron, *Rz1*, which is embedded in an alternate reading frame within *Rz*; Rz–Rz1 is thought to constitute a protein complex that somehow attacks the outer membrane. *S* encodes two proteins—the holin and the *antiholin*—as a result of translational initiation sites defined by codons 3 and 1, respectively. *R* encodes the λ endolysin, which has transglycosylase activity. The five proteins holin, endolysin, antiholin, Rz, and Rz1 represent the complete complement of lysis proteins. Here their roles in lysis and what is known about the molecular mechanisms will be described by first establishing what is known about the λ system and then comparing other less well-characterized systems to it.

The Lysis Physiology of λ

It is most revealing to focus first on the null phenotypes of the holin and endolysin (figure 10-2A). Lysis phenotypes can be reproducibly assessed by inducing a lysogen carrying a thermosensitive λ prophage under carefully controlled physiological conditions. Because of the synchronicity inherent in the sudden inactivation of the thermolabile λ repressor, *S* alleles with triggering times differing by 1–2 minutes are easily distinguishable. An even more convenient version of this, in terms of genetic manipulation, is a lysogen with a λΔ*(SR)* prophage carrying a medium copy plasmid with the promoter-proximal segment of the λ late transcriptional unit, including the promoter, p'_R, and the lysis cassette; in this system, induction of the lysis-defective prophage also causes trans-activation of the plasmid-borne lysis genes, resulting in approximately the same lysis timing as found in induction of a lysis-proficient prophage (31, 100). In figure 10-2A, key physiological properties of the λ lysis system, revealed using these kinds of assay systems on null mutants of the lysis genes, are shown:

1. S^- allows continued increase in culture mass (and, consequently, intracellular virions), whereas R^- halts growth at the normal lysis time;
2. addition of $CHCl_3$ results in instant lysis with S^-, but not R^-,
3. addition of energy poisons triggers lysis with S^+R^+, but not S^-R^+.

These observations are completely paralleled in experiments with phage T4 *t* holin and *e* endolysin mutants (119). The interpretation is that the holin triggers at a defined time to cause a lethal lesion that makes the membrane barrier permeable to the endolysin, which otherwise accumulates in the cytosol. Moreover, the holin can be artificially triggered with an energy poison, or replaced by chloroform-mediated dissolution of the membrane. Figure 10-2B shows that the triggering time is allele-specific for *S*; single-residue changes in many different positions in S can result in earlier or later lysis times (also see figure 10-1B). These are all alleles of *S105*, an *S* cistron in which codon 1 has been eliminated to ablate production of the antiholin and thus simplify the interpretation of the timing phenotypes.

The λ Endolysin and Endolysins in General

λ R is an 18 kDa soluble transglycosylase, which cleaves the glycosidic bond between *N*-acetylglucosamine and *N*-acetylmuramic acid resiues, forming a cyclic product (15). Its crystal structure reveals that it is evolutionarily related to true lysozymes, which hydrolyze the same bond (43). R appears to be dependent on the holin for crossing the membrane, although at about 3 hours after induction some R-dependent lysis of λ*Sam*-infected cells does begin to occur (R. Young, unpublished data). Whether this reflects an intrinsic membrane-penetrating ability of R or low-level access to the *sec* system is unclear. The N and C termini of R are surface-exposed in crystal structures (43), a structural feature exploited to construct fusion proteins for probing the size of the “hole” (see below).

The S105 Holin: Genetics and Biochemistry

The S holin, or S105, is a 105-residue integral cytoplasmic membrane protein with three transmembrane domains (TMDs) (52). Wild-type S protein is present at about 10^3 copies per cell at lysis (33). Nothing is known about its distribution in the membrane but, assuming 1 nm packing diameter for each helical TMD, the total amount of S can only occupy about 3000 nm^2, or 0.03%, of the membrane surface if clustered, but could span 3000 nm, approximately the cylindrical circumference of the cell, if arranged as an

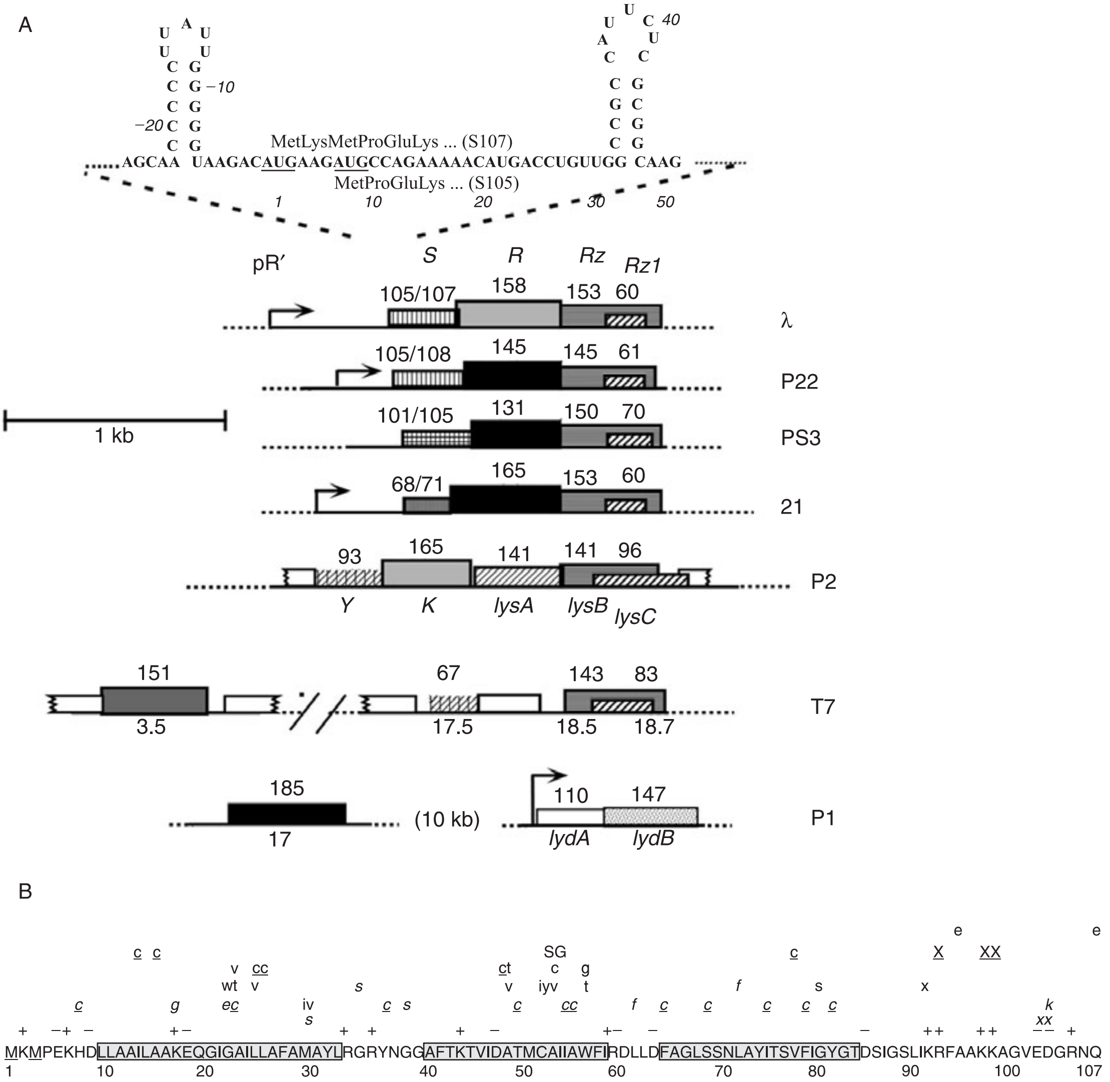

A
U A U
U U
C G
C G −10
C G
−20 C G
C G
MetLysMetProGluLys ... (S107)
U U C 40
A U
C C
C G
G C
C G
C G
AGCAA UAAGACAUGAAGAUGCCAGAAAAACAUGACCUGUUGG CAAG
MetProGluLys ... (S105)
1 10 20 30 50
pR′
S
R
Rz
Rz1
105/107
158
153
60
λ
105/108
145
145
61
P22
1 kb
101/105
131
150
70
PS3
68/71
165
153
60
21
93
165
141
141
96
P2
Y
K
lysA
lysB
lysC
151
3.5
67
17.5
143
83
18.5
18.7
T7
185
17
(10 kb)
110
147
lydA
lydB
P1
B
e e
c c SG c X XX
v cc ct c g
wt v s v iyv t f s x
c g ec iv c s c cc f c c c c c k
s xx
+ −+ − +− + + + − +− − − ++ ++ −− +
MKMPEKHDLLAAILAAKEQGIGAILLAFAMAYLRGRYNGGAFTKTVIDATMCAIIAWFIRDLLDFAGLSSNLAYITSVFIGYGTDSIGSLIKRFAAKKAGVEDGRNQ
1 10 20 30 40 50 60 70 80 90 100 107

Figure 10-1 Bacteriophage lysis genes. A: The lysis genes, with length in codons, of four lambdoid (λ, P22, PS3, 21) and three non-lambdoid (Pl, P2, T7) phages are shown: holin genes (vertical lines), endolysin genes (solid gray or black shading), nested *Rz* (horizontally lined–*Rzl* (striped) genes (96, 111), and the antiholin genes *lydB* of P1 and *lysA* of P2 (left-slanted lines) (124). The *Rz–Rzl* genes in P2 and T7 are called *lysB–lysC* and *18.5–18.7*, respectively. Identical decoration schemes for any two boxes denotes that there is detectable sequence similarity. Among the endolysins, the enzymatic activities are: black = muramidase (lysozyme); light gray = transglycosylase; dark gray = amidase. White boxes are other phage genes not involved in lysis. In the case of the lambdoid holin genes with dual-start motifs, the lengths of the holin and antiholin reading are frames, in codons, are shown above each box. The names of the λ lysis genes (*S*, *R*, *Rz*, and *Rzl*) are used for the lambdoid genes; for the *Salmonella* phages P22 and PS3, the corresponding gene names would be *13, 19, 15* and *Rzl*, respectively. Note that in T7, the endolysin gene, *3.5*, is located in an early gene cluster; its position in the lysis cassette is taken by an unrelated virion maturation gene. The positions of late gene promoters serving the lysis genes are indicated by arrows. Shown in the inset above the λ lysis cassette is the 5′ end of the *S* mRNA sequence, where two stem-loop structures control the partition of translational initiations between the start codons of the *S107* and *S105* reading frames. Nucleotide positions relative to the first base of codon l are shown in italics. Adapted from Young (120) and Johnson-Boaz et al. (65), with permission.
B: *S* mutants. The parental S sequence and characterized mutations are shown in single-letter code, with the starting residues of the S107 (Metl) and S105 (Met3) products underlined, predicted side-chain charge indicated above the sequence, and residues in the three transmembrane domains (TMDs) enclosed in gray boxes. All alleles shown are missense except that "x" denotes a nonsense allele. Keys: upper case = early lysis alleles; lower case = unconditional lysis-defective alleles; lower case italic = delayed iysis alleles; strike-through = lysis-neutral alleles. Underlined and strike-through missense alleles are all single-cysteine substitutions derived from the C51S allele. Adapted from Young (120), with permission.

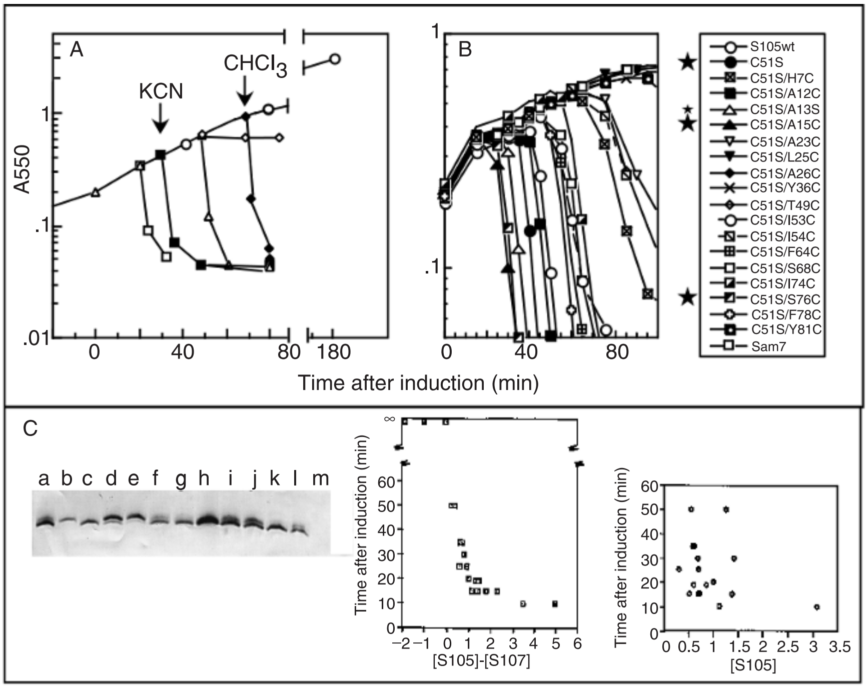


Figure 10-2 Lysis phenotypes. A: Lysis of a λ lysogen induced in exponential phase at t = 0. S^+R^+ = triangles; 5 mM KCN added to S^+R^+ = filled squares; S^+R^- = open diamonds; S^-R^+ = circles S^-R^+ = with CHC1$_3$ added, filled diamonds; $S_{a52g}R^+$ = open squares. Adapted from Wang et al. (111), with permission. B: Lysis phenotypes of single-cysteine substitution alleles. Exponential cells lysogenic for λΔ*(SR)* and carrying a plasmid with the λ lysis cassette under pR′ (also known as P_R') control were induced at t = 0. All alleles except the null *Sam7* are isogenic to the parental *S105* allele (with codon 1, the initiation site for the *S107* reading frame, inactivated). Stars indicate alleles with lysis times earlier than the *S105* parental allele. Adapted from Grandling et al. (53), with permission. C: RNA structure control of the S105/S107 partition and lysis timing partition. Left: Western blot, using anti-S antibodies, of cytoplasmic membrane fractions from cells induced for various *S* alleles with mutations in the translational initiation region, including the dominant lysis-efective *sdi* mutations that defined the upstream regulatory RNA stem-loop (see figure 10-1A, inset). The first three lanes have (a) S^+, (b) *S107* and (c) *S105*; in the latter two alleles, the AUG start codons 3 and 1, respectively, are changed to CUG, eliminating the production of S105 and S107, respectively. Alleles in other lanes: (d) *sdi3* (G$_{-8}$ to A), (e) = *sdil* (C$_{-18}$ to U), (f) G$_3$ to A, (g) inversion of *sdi* stem-loop by swapping UCCCC and GGGGG sequences, (h) UAAG (−6 to −3) replaced by GAGG, (i) C$_{33}$ to G, (j) combination of G$_{-3}$ to A and C$_{33}$ to G, (k) U$_{-24}$ to C, (l) C$_{36}$ to G and A$_{48}$ to C, and (m) both codons 1 and 3 to CUG. Mutations that weaken the *sdi* (upstream) stem-loop, strengthen the downstream loop or damage the Shine–Dalgarno sequence for *S105*—lanes (d), (e), (f), and (l)—favor S107 production over S105. Conversely, strengthening the *sdi* structure, weakening the downstream loop or improving the *S105* Shine–Dalgarno sequence—lanes (h), (i), and (k)—favor S105 production over S107. Inversion of the sdi structure separates the *S107* reading frame form its Shine–Dalgarno sequence and abolishes S107 production, and changing both start codons to CUG (m) eliminates both *S* gene products. Adapted from Chang et al. (33), with permission.

end-to-end polymer of adjacent TMDs. S105 inserts in the membrane in a *sec*-independent fashion, as judged by the insensitivity of S-mediated lysis to depletion of SecE (E. Ramanculov and R. Young, unpublished) and to shifts to the nonpermissive temperature in *secA*^{ts} and *secY*^{ts} hosts (91). Its N-terminus is blocked, presumably because it traverses the bilayer before deformylation can occur; stably solubilized in a zwitterionic detergent, it has a CD spectrum consistent with three helical TMDs (J. F. Deaton and R. Young, unpublished data).

Many *S* mutants have been isolated, including not only unconditional lysis defectives but also a plethora of alleles which accelerate or delay the onset of lysis (figure 10-1B). Thus *S* satisfies the requirement for easy genetic adjustability in terms of finding an ideal lysis time. Among the lysis-defectives, there are dominant, recessive, and "anti-dominant" (previously "early dominant") alleles; the latter are lysis-defectives which *accelerate* lysis in the presence of the wild-type allele (88)! A particularly sensitive micro-domain appears to be in the middle of TMD2 (figure 10-1B). Both λS_{A52V} and λS_{A52G} are non-plaque-formers, but for opposite reasons. S_{A52V} is unconditionally lysis-defective, whereas S_{A52G} supports catastrophically *early* lysis, at 19 minutes, before the first virion is assembled and when the level of S protein is at about 10% of its normal level at lysis (65). Similarly, at Cys51, substitution of a Ser accelerates lysis, whereas replacement with a Tyr abolishes lysis. Thus changes in bulk of only 16 or 14 Da at positions 51 and 52, respectively, cause significant premature triggering of the holin. Although less drastic, lysis timing changes are found for single substitutions all the way around the helical surfaces of all three TMDs (figure 10-2B). This collection of single-Cys mutants was generated in an effort to map the TMDs of S by cysteine-modification accessibility, starting with the cysteine-less C51S allele (52, 53). Except for two residues in the C-terminal cytoplasmic domain, every Cys substitution caused a reproducible change in lysis time. Although most of the changes caused retarded lysis, several significantly accelerated the onset of lysis, beyond that caused by the original C51S change, including substitutions in both TMD1 and TMD3 (figure 10-1B, 10-2B). Thus every potential TMD surface of S appears to be involved in setting the lysis clock. At the molecular level, S has been shown to form oligomers by crosslinking studies, but the ultimate degree of oligomerization is unknown; some lysis-defective alleles appear to be blocked at the monomer, dimer, and oligomer stages (53).

Although the sequences of the short periplasmic N-terminal domain and the cytoplasmic C-terminal domain are non-essential for lysis, the overall charge on each domain can affect timing dramatically. Especially with the C-terminus, increasing anionic or cationic character accelerates and retards the onset of lysis, respectively, and scrambling the sequence by a frameshift mutation is tolerated (18, 101). Strangely, however, although a $G_2H_6G_2$ tag inserted near the membrane interface in the C-terminal cytoplasmic domain resulted in a functional, lytic S protein which has been the source of purified holin, the simple addition of a hexahistidine tag or bulky domains such as β-galactosidase or GFP in the C-terminal cytoplasmic domain disrupt holin function (101) (R. White, A. Gruendling, and R. Young, unpublished data).

Purified S, or more properly, the purified hexahistidine-tagged variant of the S105 holin, can, at low efficiency, permeabilize liposomes loaded with a fluorescent dye, if diluted out of a chaotropic solution (99). This permeabilization fails with the S_{a52v} variant and with a *ts* version, S_{a55t}, at the nonpermissive, but not the permissive, temperature (J. F. Deaton and R. Young, unpublished data), suggesting this assay faithfully mimics the in vivo lytic event. More efficient delivery methods have been developed which should allow ultrastructural analysis of the holin-permeabilized liposomes by electron microscopy (38).

The Antiholin S107

The RNA secondary structures and other sequence elements at the beginning of the *S* gene were engineered to produce different amounts, both relative and absolute, of the holin S105 and the antiholin S107. It was found that absence of the antiholin has only a small effect in accelerating the timing, but excess antiholin effectively abolishes lysis (88). Quantitative immunoblotting of membrane samples revealed that, in general, the onset of lysis varies inversely with the excess of S105 over S107 (33) (figure 10-2C). This suggests that S107, despite being identical with S105 for all but its two N-terminal residues, is effectively a lysis inhibitor which acts by titrating out S105. This notion is supported directly by crosslinking studies (55). Depolarization of the membrane with an energy poison subverts the S107-mediated inhibition (17). The basic residue at position 2 is what distinguishes the inhibitor S107 from the holin S105; adding more basic residues at this position makes it a better inhibitor (51, 77, 101). These data suggest that the Lys2 positive charge retards the N-terminus of S107 from flipping up through the bilayer (figure 10-3), as has been demonstrated for the N-terminus of leader peptidase (29). This model is supported by the finding that fusing a signal sequence to S107 eliminates its inhibitor character, presumably because it causes *sec*-mediated export of TMD1, and converts it into a functional holin if leader peptidase activity is present to cleave off the signal (figure 10-3) (50, 51).

Despite the elegant design of the dual-start antiholin system, it is critical to emphasize that the holin S105 has an intrinsic timing function independent of the antiholin S107; the *S105* allele, in which the Met_1 start codon of S107 is inactivated, is only marginally advanced in lysis

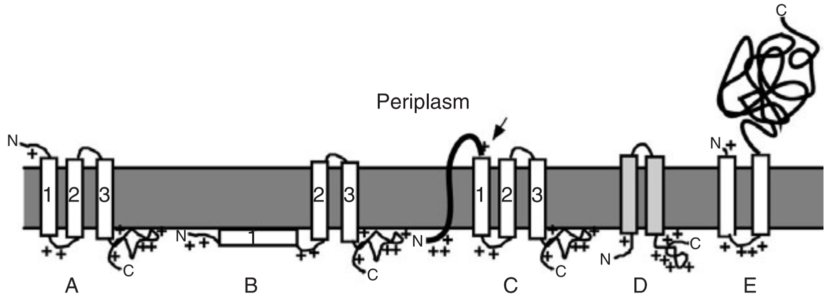


Figure 10-3 Membrane topologies of holins. A: S105. B: Putative S107 topology. C: *sec*-dependent chimeric *VIII-S* gene product, showing leader peptidase cleavage site (51). D: S^{21}. E: T4 T. Adapted from Wang et al. (111), with permission.

timing (52, 88). Models to explain the evolutionary advantages of having an antiholin are discussed below (Antiholin Diversity and Raison d'Être).

Rz and Rz1

The *Rz* (153 codons) and *Rz1* (60 codons) genes are unique in biology: one segment of DNA occupied by two genes in alternate reading frames but responsible for the same phenotype (56, 123). In the presence of millimolar concentrations of divalent cations, null alleles of *Rz* or *Rz1*, or their related P22 equivalents, cause lysis to terminate in metastable spherical cell forms (30, 122, 123). Rz1 has a signal peptidase II cleavage site and is predicted to be a 40-residue, proline-rich lipoprotein attached by its lipoyl moieties to the inner leaflet of the outer membrane, consistent with a species detected by palmitate labeling (68, 104). Rz has not been identified, but it has a consensus signal peptidase I site and presumably is a periplasmic protein. *Rz1*am in the wrong suppressor background has a dominant negative phenotype that is partially cation-independent (123).

RzRz1 equivalents can be identified in the sequences of most other phages of Gram-negative bacteria (but not for phage T4), sometimes by sequence similarity, but also, as for the P2 *lysBC* genes, from their position immediately distal to an endolysin and by the unusual nested relationship with the smaller reading frame being proline-rich and containing a signal peptidase II site (figure 10-1A) (123). *lysB*am has a cation-independent lysis delay phenotype. *lysBC*, but not *lysB* or *lysC* alone, can complement *Rz*am *Rz1*am, which is consistent with the idea that Rz and Rz1 form a functional complex (74). However, nothing is known about the function of Rz–Rz1, although it has been suggested that Rz may be responsible for the endopeptidase activity originally ascribed to R (15). There does not seem to be a link between the type of endolysin activity and the presence of *RzRz1* genes, which are found in phages with three different types of endolysins: true lysozyme (P22), amidase (T7), and transglycosylase (λ, P2) (figure 10-1A). The cation dependence of the phenotype in λ suggests that Rz–Rz1 is involved with compromising the outer membrane, or its covalent links with the murein layer. However, it is also as likely that it contributes toward more efficient endolysin activity and thus is only needed (in λ/P22 infections) when a cation-stabilized outer membrane can contribute to the residual structural integrity left after endolysin action.

Diversity in Holin–Endolysin Systems

Holin Diversity

At the time of the most recent general review on holins, there were already 105 holin genes which could be sorted into 32 families with no detectable sequence similarity (111). Now these numbers have increased to more than 250 holins in more than 50 gene families, mostly due to the acceleration of bacterial genomics and thus the parallel increase in prophage genomics. Table 10-1 has a summary, but a current and more comprehensive tabulation of holin genes and other associated lysis cistrons is available at www.thebacteriophages.org. Most can be grouped into two classes: class I, like λ S, with three TMDs, and the N-terminus out; and class II, like S^{21}, with two TMDs, with both N and C termini in the cytoplasm. The T4 holin T forms a class of its own (class III), with its relatively large C-terminal periplasmic domain, and some of the others have uncertain topologies based on primary sequence analysis (figure 10-3). Very few holins and putative holins have been subjected to functional analysis, which is difficult outside the phage genomic context because of the constraints on cloning lethal genes in multicopy plasmids. Moreover, it is clear that assessing the biological function of a protein that has as its primary function a single, saltatory, and tightly scheduled lethal event depends on reproducing the normal level of

Table 10-1 Summary of Lysis Cassette Diversity

	Gram-negative hosts		Gram-positive hosts	
A. Holin class	B. Total No. of families	C. No. of families with secretory endolysins	D. Total No. of families	E. No. of families with secretory endolysins
I	10	3	9	2
II	9	6	11	6
III	1	0	0	0
?	4	1	7	1
Total	24	10	27	9

The data are summarized from the comprehensive table deposited at the website associated with this book, which was adapted from a similar website-deposited table, with permission (112). For this table, holin gene families (columns B and D) are defined as orthologous groups with sufficient sequence similarity to permit manual alignment. If at least one holin in the family is associated with an endolysin which has been shown to support *sec*-mediated externalization or has an apparent TMD or secretory signal sequence at its N-terminus, then the entire family is counted in columns C or E. For up-to-date references, please see the website www.thebacteriophages.org expression. This is not straightforward in many bacterial backgrounds not as well characterized as *E. coli*. Moreover, a secretory or membrane protein sufficiently overproduced from a multicopy plasmid can insult the membrane sufficiently to cause release of the R endolysin (4). Also, there does not appear to be a correlation between the enzymatic activity of the endolysin and the class of holin; for example, in lambdoid phages, three different families of holins, including one class I and two class II, have been found with two enzymatically different kinds of endolysins (figure 10-1A).

In a sense, this unprecedented diversity is an obstacle for computer analysis, because there are few examples of holin families with enough members to permit comparative analysis of critical residues (see table 10-1 and website table). An exception is the class II holin family represented by the holin of the lambdoid phage 21. Alignments of the members of this family suggest that total side chain bulk is conserved at different positions in the paired TMDs (figure 10-4), as though helical packing is a critical factor, a notion also suggested by mutational analysis of the λ holin (see below).

The profusion of different holin families makes it unlikely that holins form a large, complex, well-defined structure providing a passageway through the bilayer. Instead, it suggests that it is easy to evolve a holin. Indeed, no secondary structure element is easier to evolve than a TMD; all that is required is roughly 16 residues of predominantly hydrophobic character with no net charge. This hydrophobicity will direct the protein to the membrane, and the bilayer, lacking alternative H-bonding partners for the backbone amides and carbonyls, will impose α-helical structure on the domain. What else besides two or three TMDs is required to make a holin is not clear, although the available evidence suggests the ability to oligomerize is important (see How Do Holins Work?).

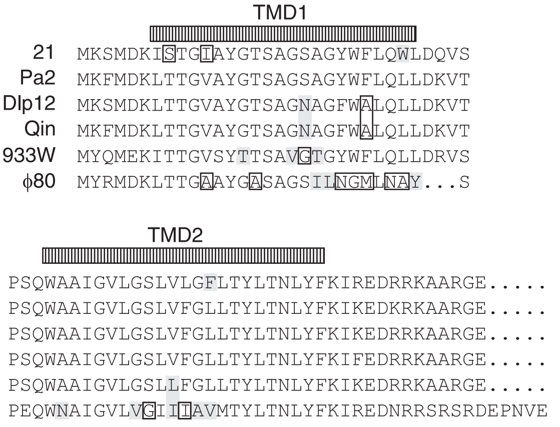

TMD1
21 MKSMDKISTGIAYGTSAGSAGYWFLQWLDQVS
Pa2 MKFMDKLTTGVAYGTSAGSAGYWFLQLLDKVT
Dlp12 MKSMDKLTTGVAYGTSAGNAGFWALQLLDKVT
Qin MKFMDKLTTGVAYGTSAGNAGFWALQLLDKVT
933W MYQMEKITTGVSYTTSAVGTGYWFLQLLDRVS
ϕ80 MYRMDKLTTGAAYGASAGSILNGMLNAY...S
TMD2
PSQWAAIGVLGSLVLGFLTYLTNLYFKIREDRRKAARGE.....
PSQWAAIGVLGSLVFGLLTYLTNLYFKIKEDKRKAARGE.....
PSQWAAIGVLGSLVFGLLTYLTNLYFKIKEDRRKAARGE.....
PSQWAAIGVLGSLVFGLLTYLTNLYFKIFEDRRKAARGE.....
PSQWAAIGVLGSLLFGLLTYLTNLYFKIKEDRRKAARGE.....
PEQWNAIGVLVGIIIAVMTYLTNLYFKIREDNRRSRSRDEPNVE

Figure 10-4 Alignment of representatives of the S^{21} class II holin family. Sequences are compared with S^{PA2}. Changes within either TMD resulting in increased or decreased side-chain bulk is indicated by shading or boxes, respectively. Dlp12 and Qin are cryptic lambdoid prophages of *E. coli* K-12 (90). PA2 is the lambdoid prophage of a porcine strain of *E. coli* (16). 933W is the lambdoid prophage carrying the Shiga-like toxin gene in the hemorrhagic *E. coli* strain 0157:H7 (86).

Antiholin Diversity and Raison d'Être

The Dual-Start Motif

As stated above, many holin genes have dual-start motifs, with two potential start codons separated by at least one codon for a basic amino acid. In the few other cases tested (P22 *13*, 21 *S*, and ϕ29 *14*), eliminating codon 1 *accelerates* the onset of lysis, which is presumptive evidence that the dual start functions analogously to that of *S* (6, 77, 102). At least for S^{21}, the holin gene of lambdoid phage 21, the longer product has been shown to have antiholin character, based on the N-proximal basic residue (6). In many of these cases, however, including S^{21}, the holins are class II sequences (only two TMDs). Thus the molecular basis of the inhibitory function must be fundamentally different because the N-terminal domain of a class II holin is not available for transiting the bilayer (figure 10-3).

Irrespective of the basis of the inhibitory character of the longer gene product of dual-start holin genes, at least in the two cases examined to date, it is clear that RNA secondary structure plays a key role in regulating the absolute and relative rates of initiation at the two start codons. In λ *S*, two stem-loop structures, one overlapping the strong consensus Shine–Dalgarno sequence for *S* and a second spanning codons 11–16, are involved in achieving the approximately 1:2 partition of initiations at codons 1 and 3 (figure 10-1A). Weakening the upstream structure or strengthening the downstream structure leads to dominant lysis-defective *S* alleles producing an excess of S107; conversely, strengthening the upstream structure or weakening the downstream structure increases the S105:S107 partition and accelerates the onset of lysis (21, 33, 77). A 60 kDa host protein binds the RNA immediately adjacent to the upstream stem-loop. Absence of this upstream sequence greatly reduces the translational efficiency of the *S* cistron, indicating that the host factor and the RNA may play a role in regulating *S* and also perhaps in the S105:S107 partition (32, 76).

In S^{21}, there is only an upstream stem-loop, and although it only partly occludes the Shine–Dalgarno sequence, weakening it again produces a partially dominant lysis-defective allele that produces only $S^{21}71$, the antiholin (6). However, elimination of codon 1 greatly reduces initiations at codon 4, the start codon for $S^{21}68$, the holin, and thus masks the early lysis phenotype expected from loss of the antiholin. Mutagenesis of the translational initiation region suggests that the partition of antiholin versus holin starts in S^{21} results from initiation complexes forming over codon 1 and then frequently diffusing downstream to codon 4 (6).

Why Dual-Start Antiholins?

Why do so many holin genes have a dual-start motif, and thus, presumably, encode an antiholin? In λ, the short lysis delay associated with the normal level of S107 antiholin production could easily be achieved more simply by changing a residue in the holin S105. First, of course, it seems likely that there may be physiological conditions where the partition of antiholin–holin starts shifts in favor of the former, and thus results in a very significant delay, if not complete blockage, in lysis. A quantitative model for lysis timing suggests that a reduction in the productive capacity of the host or in the availability of fresh hosts in the medium should lead to a selection for longer lysis times (110). The dual-start system could provide a phenotypic adaptation to such changed conditions, if it were programmed to detect the appropriate physiological signal and effect an inversion of the normal antiholin–holin ratio.

Another perspective is that the S107–S105 system contributes to the saltatory nature of holin-mediated lysis. The wild-type *S* gene produces about twice as much S105 as S107. Because S105 preferentially dimerizes with S107, half the S105 is sequestered in heterodimers, resulting in two thirds of the total S protein in nonfunctional, or less functional, heterodimers (55). Whatever constitutes the spontaneous triggering event would collapse the energized state of the membrane, thereby converting the nonfunctional heterodimers into functional holin dimers (figure 10-5). Thus at the instant of triggering, the functional holin level is suddenly tripled, presumably expediting the disruption of the membrane barrier. This may be the best way to achieve a reasonable time delay, by reducing the rate of accumulation of functional holin, but also ensuring that the sufficient holin can be recruited at the instant of triggering to effect rapid release or activation of the endolysin.

rI and Lysis Inhibition in T4

Not all antiholins are alternate-start products of the same gene encoding the holin. The best studied independent antiholin is the RI protein of phage T4. Its gene, *rI*, was one of the four T4 plaque-morphology genes (the others being *rIIA*, *rIIB*, and *rIII*) identified by Hershey in experiments that are regarded as among the founding works of modern molecular genetics (58). The phenomenon underlying these phenotypes is called LIN, for "lysis inhibition." In a T4 infection, the *imm* gene is expressed early; Imm is a cytoplasmic membrane protein which blocks entry of the genome of a T4 phage attempting to superinfect the infected host. Somehow, this aborted secondary infection leads to imposition of LIN, in which the normal scheduled lysis at 25 minutes after infection is blocked. LIN can persist indefinitely as long as secondary infection events continue to occur. The LIN state in every respect mimics the physiology of a holin-defective infection, in that intracellular virions and endolysin accumulate far beyond normal levels, except that the LIN collapses if the energized state of the membrane is subverted by an energy poison (2).

An Ancient Player Unmasked: RI Is an Antiholin

Despite its ancient and well-characterized physiology, and its historical importance, the mechanism of LIN was elusive for more than five decades, mainly because of the difficulties of manipulating the T4 infective cycle and the bewildering number of *r* loci identified over the years since Hershey's initial work (2). Recently, analysis of this extensive literature led Paddison et al. (83) to conclude that among the original *r* loci only *rI* is required for LIN in all bacterial hosts. Moreover, another LIN-defect locus, *rV*, was shown to be allelic to *t*, suggesting that the *r* genes were

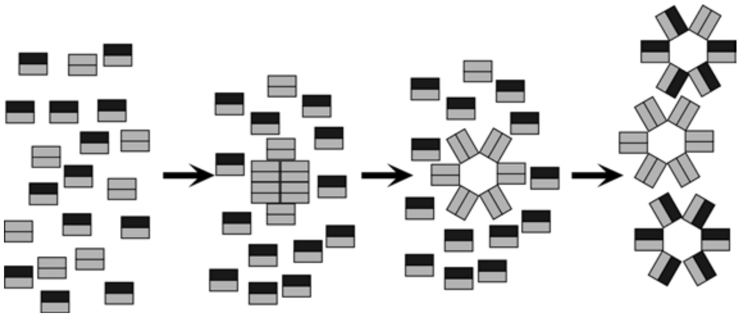


Figure 10-5 Holin function and the dual-start motif. Holins and antiholins from dual-start holin genes are depicted as light and dark shaded rectangles, respectively. The cartoon reflects the normal 2:1 proportion of the holin S105 and S107 and assumes that dimerization is the first step in hole formation. According to this model, each S107, which preferentially heterodimerizes with S105 (55), removes one S105 from the pool of holin monomers involved in the timing of hole formation. At the instant of the first hole formation, the collapse of the membrane potential allows the S107 protein to assume the conformation of the S105 holin. Thus, the scheduled triggering results in tripling the amount of functional holin homodimers available for hole formation. Adapted from Gründling et al. (55), with permission.

a signaling system targeting the holin (42). Finally, it has been shown that LIN can be imposed upon the chimera λ*t*, in which *S* is replaced by *t*, by superinfecting with T4 (89). LIN is not imposed upon the isogenic λS^{+} when it is superinfected with T4. The imposition of the LIN state was found to depend on the allelic state of the *rI* and *t* genes of the superinfecting T4 phage, indicating that the LIN-sensitive state of the T holin is transient and suggesting that the antiholin acts by impeding entry of the holin into oligomeric “pre-hole” assemblies (89). Remarkably, using the same λ*t* system to test for LIN, the requirement for superinfection by T4 can be replaced by simple induction of a plasmid-borne copy of *rI* (98a). Moreover, crosslinking in vivo with formaldehyde revealed that the antiholin RI binds to the membrane-embedded holin. Thus, at last, the outlines of a molecular explanation for the principal sacrament of the Phage Church are apparent, but many questions remain. It is unclear why simple expression of *rI* from a plasmid eliminates the need for secondary T4 infection. The key to this may lie in the instability of RI, which has only been detected after cross-linking to T. Perhaps the ectopic periplasmic localization of the 169 kb genome of the superinfecting T4 particle, or proteins ejected from its capsid, somehow stabilize RI, which appears to consist of a 6 kDa periplasmic domain tethered to the membrane by an N-terminal TMD (83). It is unknown how RI recognizes T, and how its binding prevents T from entering the pathway to lysis triggering. The roles of *rIIA*, *rIIB*, and *rIII* in LIN are also mysterious. Although none is essential in *E. coli* K-12 hosts, each has a dramatic LIN-defective phenotype in some backgrounds. Analysis of lysis phenotypes suggests that RIII, which is predicted to be a cytoplasmic protein, may stabilize the RI–T interaction (89). The ability to reconstruct LIN apart from the complexity of the T4 vegetative cycle promises that rapid progress in this venerable phenomenon may be at hand.

Other Independent Antiholin Genes

Genetic or physiological evidence for antiholin genes is available for the classical coliphages P1 and P2. In P1, although the known lysis genes are all under control of the late gene activator, the holin and antiholin genes, *lydAB*, are adjacent but unlinked to the endolysin gene, *lyz* (formerly *17*), encoding a T4 E orthologue (figure 10-1A). Remarkably, the antiholin LydB is essential to obtain a productive burst; in its absence, lysis occurs catastrophically early (108). LydB is tethered to the membrane by a single N-terminal TMD, with a highly hydrophilic domain of about 14 kDa in the cytoplasm. Experiments with a λΔ*S::lydA* chimera have shown that LydB retards lysis mediated by LydA, a class I holin, but not S-mediated lysis, in the context of the λ vegetative cycle (M. Xu, D. K. Struck and R. Young, unpublished data). In P2, the lysis cassette is comprised of genes *Y K lysA lysB lysC* (124). Y, K and LysB/LysC are, respectively, a class I holin, an orthologue of the λ R endolysin, and, as noted above, functional analogues of λ Rz/Rz1. A null mutation in *lysA* causes a 5 minute acceleration of the onset of lysis, and overproduction of LysA in trans to an induced λ chimera, with *Y* replacing *S*, but not with induced λ, causes a significant lysis delay (T. Park and R. Young, unpublished data), suggesting that LysA is a Y-specific antiholin. LysA appears to be an integral membrane protein with four TMDs, and is predicted to have two basic residues at its cytoplasmic N-terminus, reminiscent of the antiholins of the class II holin genes (124). However, the molecular basis of the LysA-mediated lysis delay and the LysA-Y specificity is not known.

Endolysin Diversity

Four different enzyme activities have been associated with endolysins: “true lysozyme” (also known as muramidase; e.g., T4 *e* lysozyme, P22 gp*13* lysozyme), which hydrolyzes the 1,4 glycosidic bonds in the murein; transglycosylase (e.g., λ R or P2 K), which attacks the same bonds but forms a cyclic 1,6-anhydro-*N*-acetylmuramic acid product; amidase (e.g., T7 gp*3.5*), which hydrolyzes the amide bond between the *N*-aceytlmuramic acid and L-alanine residues in the oligopeptide crosslinking chains; and endopeptidase (e.g., ϕ11 lysin), which attacks the peptide bonds in the same chains (79, 119, 121). In some cases, more than one of these activities is found in the same protein (79). Again, whenever the notion has been tested, with important exceptions noted below (The Other Paradigm: *sec*-Mediated Lysis), endolysins of different enzymatic types complement each other. Nonspecificity with respect to holins is clearly seen by comparing λ and P22, which have nearly identical class I holins but completely different endolysins, and also P22 and PS34, which have nearly identical lysozymes orthologous to T4 E but unrelated class I and class II holins, respectively (figure 10-1A). Almost all endolysins are late proteins; a notable exception is T7 gp*3.5*, which is an early protein and has an important early function, as specific inhibitor of T7 RNA polymerase (34, 75). The holin–endolysin–RzRz1 lysis gene cassette conserved at the start of the late gene transcript of all lambdoid phages is present in the late T7 transcript, except that the endolysin gene is replaced by an unrelated gene involved in phage DNA maturation (figure 10-1A).

Many endolysins of phages specific for Gram-positive hosts have a modular structure, with a C-terminal domain determining binding specificity exquisitely sensitive to the differences in the structure of the cell wall (37, 40, 41, 98). For example, the Cpl1 endolysin, a T4 lysozyme orthologue, from the pneumococcal phage Cp1 has a C-terminal domain specific for the choline component of the teichoic acid of its host, whereas the lysozyme from phage Cpl7, with a similar enzymatic domain, is choline-independent. Similarly, the Dp1 endolysin, an amidase, has a C-terminal cell wall binding domain orthologous to the choline-dependent domain of Cpl1. The specificity and lytic activity have been exploited in the use of a phage endolysin as an aerosol to ablate pneumococcal infections in the mouse nasopharyngeal cavity, a development that highlights the possible use of phage proteins as antibacterial factors (80).

Emerging Perspectives and Questions about Holin–Endolysin Lysis

The Other Paradigm: sec-Mediated Lysis

A Dogma and Its Death

Although the first comprehensive review of phage lysis did not appear until 1992, the clarity and elegance of the λ paradigm rapidly imposed itself as a kind of dogma. The complementary phenotypes of two *S* missense alleles, S_{A52G} and S_{A52V}, with unconditional plaque-formation defects illustrate this point. The former causes lysis too early, before the first virion is assembled, whereas the latter is lysis-defective, indistinguishable from the null *Sam7* in terms of the unabated accumulation of intracellular endolysin and virions (53, 88). This is formal genetic proof that, for λ, the holin has two essential functions: *to disrupt the membrane to allow escape of the endolysin to its murein substrate, and to effect the scheduled termination of the infective cycle*. In some phage systems, however, holin genes are not essential for plaque formation; for example, T7*17.5*am mutants make plaques. Examination of this problematical phenotype emphasizes the point made above, that holin nulls *increase* virion production. Presumably, when T7*17.5*am virions are plated on a lawn of sensitive cells, the lack of timely lysis in each infection is compensated by a greatly increased accumulation of virions per infection, such that when the slower and nonspecific insults to the general health and envelope integrity of the infected host associated with the aggressive T7 vegetative cycle finally lead to membrane leakiness and lysis, the net result is formation of a plaque of substantial size. This perspective explains why holin genes had been identified as essential genes only in lambdoid and T-even phage, where fortunately the nonspecific "rotting" of the infected cell at times long past the normal onset of lysis does not result in a plaque.

Just as the λ lysis paradigm was becoming clarified, at least in terms of its molecular components, a paradigm shift began with studies by Santos and colleagues on the phage fOg44, which grows on the Gram-positive bacterium *Oenococcus oeni* (85, 95). In this phage, the predicted product of the *lys44* gene is a muramidase with an N-terminal signal sequence; moreover, Lys44 is processed by leader peptidase during infection (95), presumptive evidence that the endolysin is exported by the *sec* machinery. Nonetheless, fOg44 apparently has a holin gene, *hol44* (85), and its co-expression with *lys44* in the heterologous *E. coli* environment leads to more efficient lysis (95). A survey of orthologous endolysins from other phages of Gram-positive bacteria suggested that some, but not all, of these endolysins had N-terminal sequences resembling secretory signals, although in every case an adjacent holin gene orthologous to the putative *hol44* was also present. Unfortunately, definitive physiological experiments with these systems to ascertain holin-negative phenotypes have not yet been practical. Nevertheless, it is clear that in these phages the holin is not required for export of the endolysin. The authors proposed that the exported, processed endolysin is either inactive, or less active, until the holin exerts its timed disruption of the membrane. According to this scheme, holin triggering would, at minimum, collapse the energized state of the membrane and activate the pre-localized endolysin. That this may be the mechanism of activation of the exported endolysin is suggested by the fact that, in Gram-positive bacteria, host muralytic enzymes are present in an inactive state in the cell wall and can be activated by energy poisons or anoxic conditions. Recent studies with pH-sensitive fluorescent dyes and murein modification reagents have indicated that the murein layer in Gram-positive bacteria is maintained at a much lower pH than the ambient medium, at least in respiring cells (27, 28, 66). Some endolysins have high pH optima, so the energization of the membrane may be directly responsible for maintaining the secreted endolysins in an inactive or less active state.

Echos in Coliphage: Secreted Endolysins

Sequence comprisons also reveal that a number of endolysins from phages of Gram-negative bacteria also have N-terminal sequences that could engage the *sec* system. There is a hydrophobic N-terminal extension on the predicted endolysins, all orthologs of T4 E, from coliphage P1 and from the lambdoid phages 21, P22, PS119 and PS3 (figure 10-6). The length of the hydrophobic sequence and the distribution of flanking charged residues are, however, indicative of a N-terminal TMD, a "signal anchor" or "uncleaved signal sequence," rather than of a cleavable signal peptide like the endolysin in fOg44. Experiments with clones of P1 *lyz* confirm that lysis, albeit somewhat delayed and less saltatory, follows after induction of this endolysin gene in the absence of a holin gene (118b). Moreover, cyanide treatment triggers early lysis and dramatically accelerates lysis already begun, whereas 1 mM azide, a specific inhibitor of SecA, blocks lysis. A detailed analysis of P1 Lyz confirmed that it is initially localized as an N-terminally tethered, enzymatically inactive form, and then is activated by release from the membrane (118b). This release occurs spontaneously at a low rate, thus accounting for the holin-independent lysis supported by *lyz* clone, or quantitatively, when the membrane is depolarized artificially or by the triggering of the P1 holin. The N-terminal TMD of Lyz was designated as a SAR

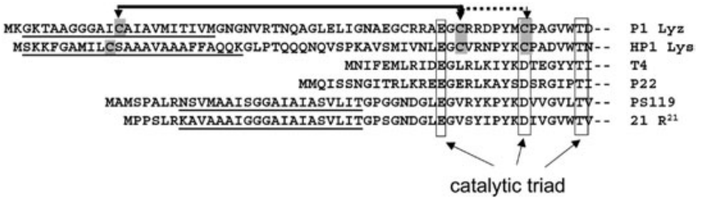

MKGKTAAGGGAICAIAVMITIVMGNGNVRTNQAGLELIGNAEGCRRAEGCRRDPYMCPAGVWTD-- P1 Lyz
MSKKFGAMILCSAAAVAAAFFAQQKGLPTQQQNQVSPKAVSMIVNLEGCVRNPYKCPADVWTN-- HP1 Lys
MNIFEMLRIDEGLRLKIYKDTEGYYTI-- T4
MMQISSNGITRLKREEGERLKAYSDSRGIPTI-- P22
MAMSPALRNSVMAAISGGAIAIASVLITGPGGNDGLEGVRYKPYKDVVGVLTV-- PS119
MPPSLRKAVAAAIGGGAIAIASVLITGPSGNDGLEGVSYIPYKDIVGVWTV-- 21 R21
catalytic triad

Figure 10-6 Signal-anchor-release (SAR) domains in endolysins of Gramnegative bacteria. The N-terminal sequence of the endolysins of the lambdoid phages 21 (22), and PS119 (GenBank accession CAA09710), P22 (30), and the non-lambdoid phages P1 (30), T4 (81), and HP1 (7) are shown. The N-terminal SAR domain is underlined in the P1, HP1, and R^{21} endolysins (118b). The catalytic triad correspnding to residues E11, D20, and T26 of T4 lysozme are boxed (87). The dashed brackets arrow shows the disulfide linkage that inactivates membrane-tethered P1 Lyz by occupying the catalytic Cys residue, and the solid brackets arrow shows the disulfide linkage in the active, periplasmic form of the enyzme (118a).

(signal-anchor-release) domain because of its unprecedented ability to escape from the bilayer. Further experiments revealed that the membrane-tethered form of Lyz was inactive because of disulfide bond occupying an active-site Cys residue, and that activation of the muralytic activity occurs when the free Cys in SAR domain, once liberated from the membrane, triggers a disulfide bond isomerization (118a). Structural analysis revealed that the isomerization is accompanied by a drastic conformational change that completely reorganizes tha catalytic cleft. Thus P1 Lyz is regulated by topological, covalent, and conformational constraints, in addition to being subject to the timing schedule of the P1 holin–antiholin system. Other endolysins, like R^{21} (figure 10-6) have SAR domains but do not have either a Cys residue in the membrane-embedded sequences or a Cys at the active site and thus must have an alternative way of regulating the exported muralytic activity.

How Do Holins Work?

The foregoing has promoted the idea that holins are remarkable proteins, honed by powerful evolutionary forces to be the simplest and most adjustable of, if not clocks, biological timers. Moreover, it also seems clear that holins are sensitive to the energy state of the membrane and thus, perhaps, constitute the simplest of gated membrane assemblies. The reader has no doubt noticed, however, that there is very little mechanistic insight into how holins work, however elegant and amusing the phenomenology. The two interesting, and unanswered, questions are: how do holins disrupt the membrane, and how do they accomplish this disruption according to a strict temporal program?

How Big Is the "Hole"?

Nothing is known about the nature of the membrane lesions. One experiment has been done to calibrate the scale of the "S-hole." Fusions of R were created with full-length *lacZ*, creating a gene encoding a hybrid R-βgal protein of 1189 residues (109). The fusion proved to be fully active and tetrameric, with a mass of nearly 0.5 MDa; remarkably, it is also fully active in lysis, with only a short delay after the holin triggering, presumably due to slower diffusion of the huge chimera in the periplasm. This result suggests that at least some of the lesions created by the triggering of S are of a size comparable to the large-scale lesions, in excess of 30 nm diameter, created by the thiol cytolysins such as streptolysin O (84, 97). Preliminary electron micrography, using purified S to permeabilize liposomes, suggests that artificial membranes, at least, are completely disrupted by the holin, rather than permeabilized by the assembly of a regular pore structure (J. F. Deaton, C. Savva, J. Sun, A. Holzenburg and R. Young, unpublished data).

Are All Holes, and Holins, Equal?

If some holins are required only to effect a scheduled collapse of the membrane potential, thus activating presecreted endolysins, can they function to allow the fully folded, secretion-less endolysins such as λ R or T4 E to attack the cell wall? Definitive experiments appropriately controlled for physiological levels of expression are not yet available. However, recent experiments have shown that premature triggering of the S holin with an energy poison allows lysis with R but not with a full-sized R-β galactosidase, suggesting that the potential "hole size" increases with the accumulation of the holin in the membrane (109). An attractive, unifying notion is that all holins produce a range of lesion sizes, perhaps dependent on the population of different n-mer states in the membrane at the time of triggering, which, in turn, may be kinetically defined, rather than in an equilibrium state. The size of the lesion may simply reflect the size of pre-lesion aggregates in the membrane, which, in turn, could be related to the mechanism of spontaneous triggering and its temporal dependence. That is, significant membrane disruption, and thus the possibility of the release of pre-folded endolysins, may be an inevitable consequence of the scheduled triggering mechanism, rather than its main goal.

How Do Holins Keep Time?

Ultimately, the most intriguing issue remains how holins function as such precise timers. Our understanding of this phenomenon is still at the stage of learning which questions can be answered. The *S* gene remains the paradigm. Two recent experiments help constrain the range of models. A common characteristic of all holins tested to date is the triggerability by energy poisons. One of the first models proposed for the timing capacity of the λ holin had it that the spontaneous lysis mediated by S comes about because, as S accumulates in the membrane, it causes ion leakage and gradually titrates out the membrane potential. When the leakage became unsupportable and the energization of the membrane collapsed, hole formation would immediately ensue, as though an energy poison had been added externally. This model was falsified by experiments using the system developed by H. Berg and colleagues to study the powering of the bacterial flagellum by the proton-motive force (pmf) (46). In these experiments, bacteria were tethered to a glass slide by antiflagellin antibodies and then their rotation speed monitored by videomicroscopy; under appropriate conditions, the rotation speed of the tethered cell is proportional to the pmf over a wide range of voltages. By inducing a cloned λ lysis gene cassette in the tethered cell, it was shown that the rate of cellular rotation is maintained up until the instant of triggering; at this point, rotation is halted and, within seconds, the cells burst (figure 10-7) (54). Thus, holins kill without warning; the hole-formation pathway does not involve intermediates that affect the pmf or ion gradients at all. Moreover, by titrating the pmf with the uncoupler dinitrophenol, it could be shown that S triggered as soon as the pmf was reduced about 40%, which would be at about −100 mV, assuming the normal pmf is −180 mV (54). Thus S is not only a simple biological timer but also one of the simplest "gated" membrane proteins, sensitive to changes of about 60 mV.

A simple alternative to the leakage model which may explain the timing of the saltatory hole formation is the "critical concentration" model. This model simply states that S accumulates in the membrane until a critical two-dimensional concentration or number density or mole fraction is achieved. At this point, the hole structure begins to form. If the new hole structure collapses the membrane potential, then the entire population of S molecules is triggered. A precedent for this is known in the formation of the two-dimensional crystalline array of bacteriorhodopsin in the purple membranes of halobacteria; bacteriorhodopsin monomers accumulate in the membrane until the critical concentration is reached, after which the crystal form accumulates and the monomer population stays constant (62, 63). There are more sophisticated versions of this kind of model, depending on how many n-mer substrates there are before the hole-forming oligomer is formed. A simple prediction of this model is that if there are two different, non-interacting holins accumulating in the same cell, then the holin reaching its critical concentration first will dominate the timing of lysis; the presence of a holin which will reach its critical concentration later should be irrelevant. A preliminary test of this model has been done, using λ *S* and T4 *t* mounted in prophage and trans-activatable plasmids under pR′ control. Ignoring for the moment the unproven assumption that S and T do not interact, the results were unambiguous: lysis triggers much faster with both holin genes being expressed than with either of them expressed under isogenic conditions (89). Simply put, it seems that both holins were doing something to the cell, before lysis, and the effects of the two holins were additive. Several reservations must be attached to this conclusion, however. First, it is of course possible that the two holins do interact, despite the absolute lack of sequence similarity, or interact with a common protein target in the membrane. The failure to find any bacterial mutations that affect holin function, as opposed to holin accumulation, may simply be bad luck. Also, it has not been shown that the kill-without-warning feature of S is shared by T, so the early lysis might reflect T-mediated leakage leading to early triggering of S.

In any case, with the perspective of the tethered cell experiments, the timing problem is even more sharply

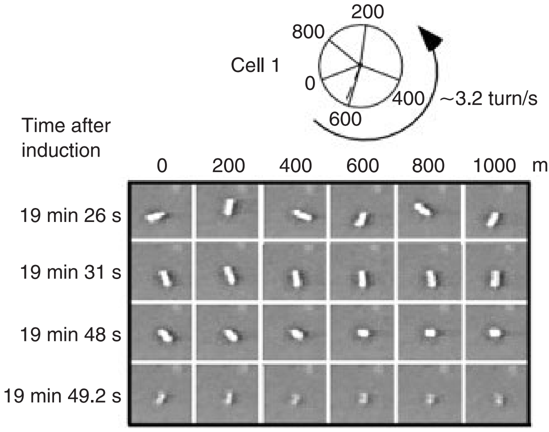


Figure 10-7 The phage λ holin kills without warning. At t = 0, cells tethered to a microscope slide by antiflagellin antibodies were induced for the expression of the plasmid pSI05$_{A52G}$, which carries the λ lysis cassette, including the S_{A52G} early lysis allele, under the control of its cognate promoter, pR′. Single frames chosen from the recordings of a representative cell are depicted here to illustrate the saltatory nature of holin killing. Starting from the time point indicated to the left of the panel, single frames were captured every 200 ms. After induction of the lysis genes, the tethered cells rotate at high and constant speed (first row). About 20 min after induction, rotation of the cell abruptly slows and stops completely within 1–3 (second row). Cell lysis, due to digestion of the cell wall by the λ R endolysin, occurs within several seconds after the sudden stop in rotation (third and fourth rows). Adapted from Gründling et al. (54), with permission.

defined. From an evolutionary viewpoint, these results are not surprising. S functions, as presumably all holins do, to terminate the infective cycle at an optimal time, consistent with the environment and the condition of the host. Sudden triggering and release or activation of excess endolysin without prior insult to the membrane serves this purpose perfectly. The capacity for macromolecular synthesis is undisturbed up until the instant of lysis, thus maximizing the burst size for the particular length of vegetative cycle and reducing the dwell time of finished virions in the nonproductive corpse.

"Death Rafts": A Model for Holin Function

The extensive genetic analysis of *S* has provided clues to thinking about the timing issue:

1. S105 alone exerts a precise timing regimen, so the S107 antiholin is not directly involved in the core temporal regulation.
2. S105 protein deleted for most of the highly charged cytoplasmic C-terminus retains a sharply defined lysis schedule. Thus, although increasing the balance of cationic charge in the cytoplasmic domain can retard the triggering time, this domain is not required for the intrinsic timing function (18). Similar analyses at the N-terminus are more difficult to interpret, because increasing the number of positively charged residues in this domain would in effect re-create a S107-equivalent antiholin protein (101).
3. Mutations mapped to both of the interhelical loops and to all the helical surfaces of all three TMDs change the timing of lysis unpredictably, apparently at the level of function rather than synthesis or stability (52, 53).
4. Mutant alleles defective in lysis have unpredictable phenotypes in the presence of the parental allele. Some are dominant, some are recessive, and some are "antidominant"; that is, some lysis-defectives, in the presence of S^+, advance the triggering time as much as a second parental allele (88).
5. Lysis-defective alleles appear to be blocked either at the monomer–dimer, dimer–oligomer, and oligomer-"hole" transitions (53).

It is difficult to rationalize these findings with any model where a defined, regular structure is assembled in the membrane. Moreover, the unprecedented diversity in holin sequences and structures seems unlikely to reflect many different ways to make a sophisticated pore structure. The common functionality of all these proteins, differing not only in amino acid sequence but in membrane topology, suggests that evolving into a holin sequence must be relatively easy. Taken together, the phenotypic complexity and genetic diversity suggest that there must be a common fundamental property of S and other holins. It is proposed here that this property is that holins are capable of accumulating, by intimate packing of their TMDs, in oligomeric patches or rafts in the membrane in which the lipid is largely, if not completely, excluded from the protein–protein interfaces (figure 10-8). As these patches grow, either by merging of smaller rafts or by accretion of dimers into rafts, the forces on these patches change, dominated early by lipid–protein interactions but, as the average patch diameter increases, later by protein-protein interactions. The local surface tension, conductance, and capacitance properties of these patches are obviously very different from that of the normal membrane, especially considering the nearly megavolt per centimeter voltage across the bilayer. Moreover, these properties would be expected to be influenced dramatically by every change in every surface and loop of the S protein. The simplest model is that at some patch size, a portion of a raft momentarily gives way to dispersive forces, creating a local, microscopic ionic current and local depolarization. With functional *S* alleles, this would lead to a concerted local conformational change (perhaps in the average angles from the orthogonal of the TMDs, for example) that shifted the balance in favor of dispersive forces. Thus an acute local collapse of the energized state of the membrane

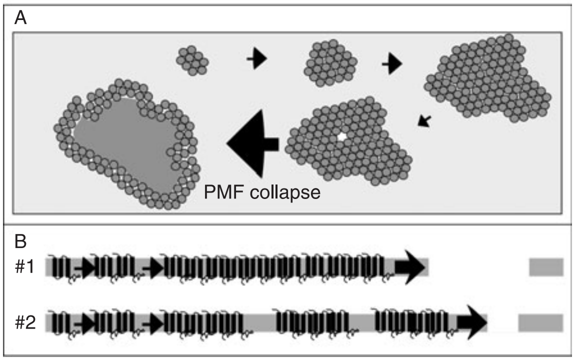


Figure 10-8 The "death raft" model for holin function. A: Holin rafts. It is proposed that holins accumulate in rafts in the membrane, shown here in a schematic top-down view, and that intimate intermolecular and intramolecular helix-packing between the TMDs of the holins largely excludes lipids. Each circle represents a single holin molecule. Spontaneous formation of an aqueous channel by thermal fluctuation is depicted. The localized depolarization causes a conformational change in the holins leading to asymmetric disruption of the helix packing, exposure of a relatively hydrophilic surface, and dispersion of the subunits into the holin lesion. B: "Hole" size. Based on the raft model above, a rationale for the formation of different-sized lesions by holins, depicted in a cross-sectional schematic of the membrane. Large lesions would be formed by the triggering of the large rafts of holin #1, whereas smaller lesions are formed in #2, which could represent a mutant of holin #1 or a heterologous holin that normally functions with a *sec*-exported endolysm and thus is not required to form a large lesion to allow endolysin release. Adapted from Wang et al. (109), with permission.

would be propagated, rapidly triggering the collapse of all the rafts in the membrane. The terminal phenotype might be ragged holin-lined lesions in the membrane, perhaps reflecting the distribution of raft sizes at the instant of triggering.

To take this kind of model seriously would require, at minimum, direct demonstration of large, two-dimensional arrays of S, a task made difficult by the extreme triggering-sensitivity of cells in which functional S is accumulating, where even brief anoxia imposed on the cells causes instant lysis. Also, if energized liposomes can be loaded with S protein labeled with fluorescent probes, changes in the physical state of S associated with sudden depolarization might be correlated with the extent and kinetics of membrane disruption, assessed by dye release and visualization with electron microscopy.

Why Are There Antiholins?

The existence of S107 as an antiholin for the λ holin at first seems an elegant solution to the timing problem, until one contemplates the very minor phenotype of ablating the antiholin. As noted above, under laboratory induction conditions, lysis timing is accelerated only about 5 minutes in mutants producing only S105 at its normal rate (88). It would be simple to find a missense allele of *S* with the slightly early lysis time, so why have the antiholin at all? In the P1 system, the timing of lysis supported by the class I holin LydA in the absence of the antiholin LydB is so early that no virions have been assembled. However, even here, extrapolating from the canonical class I holin S, it should be easy to select a *lydA* mutant which triggers at the normal time, and thus eliminates the need for an antiholin. The answer may lie in the remarkable properties of the T4 antiholin RI, which under certain circumstances is activated to override the normal timing schedule of the holin T. In this case, the antiholin provides a real-time adjustment for the evolutionarily selected holin timing schedule. One may suppose that there are conditions where the S107/S105 ratio is altered and thus the normal timing of the parental S105 protein is overridden. In λ, RNA structures define the proportion of holin and antiholin, and, as noted above, there is evidence for an RNA-binding host factor being involved in the translatability of the *S* mRNA. There may be environmental conditions which, through this host factor, determine the holin–antiholin production ratio and thus directly override the lysis timing programmed into the S105 sequence. Similar structures are found in other dual-start motifs (22).

Lysis Without Lysozyme

It appears that all phages with double-stranded nucleic acid genomes use the holin–endolysin system for lysis. This must reflect a powerful evolutionary pressure for optimization of the lytic event, both in terms of efficiency of release of the progeny virions and, probably more so, in terms of terminating the vegetative cycle at a time appropriate for the particular host and environment. However, no lysozyme is encoded or produced by three classes of very simple lytic phages with small, single-stranded nucleic acid genomes: the ssDNA *Microviridae* and the ssRNA male-specific *Leviviridae* and *Alloleviridae*, the prototypes for which are ϕX174, MS2 and Qβ, respectively (figure 10-9). It has been a mystery for decades how these small viruses achieve lysis of the host and liberation of the progeny without producing a muralytic activity. In each case, a single phage gene was shown to be necessary and sufficient for host lysis: *E* for ϕX174, *L* for MS2 and A_2 for Qβ (5, 8, 60, 114).

Probably the two most heavily promoted models for lysis in these "single-gene" systems have been:

1. *Autolysis*: according to this view, the viruses induce endogenous muralytic activities which, although functional in vivo, are not detectable by in vitro assays (19, 71, 72, 107).
2. *Transmembrane tunnels*: according to this model, promulgated for ϕX174, the lysis protein E forms an oligomeric tunnel structure through the entire envelope, allowing the cytoplasmic contents to escape (116–118).

Recently, the molecular basis of the lytic function of the ϕX174 E protein and the Qβ A_2 protein has been established unequivocably, by a combination of genetic and biochemical techniques. In both cases, the lysis protein has been shown to be an inhibitor of a specific enzyme in the pathway for cell wall synthesis (10, 12, 14). This may be regarded as the fundamental mechanism, although the sequelae of inhibition of cell wall synthesis, either by phage lysis protein or by fungal antibiotic, may very well be considered "autolysis," in the sense that some host proteins are involved in the terminal phenotype. However, the transmembrane tunnel model is definitely ruled out, since E has been shown to be a specific enzyme inhibitor rather than a component of a pore structure spanning the envelope.

ϕX174 E

The ϕX174 genome has only 10 genes, and the lysis gene, *E*, spans only 91 codons in an alternate reading frame embedded in an essential gene *D* encoding a capsid scaffolding protein (figure 10-9). E is a membrane protein; its first 35 residues define the lytic function and the membrane insertion domain. Although its topology is uncertain, several lines of evidence indicate the C-terminus is cytoplasmic. Replacement of the entire C-terminal cytoplasmic domain by β-galactosidase, GFP, or any folded domain has no effect on lysis of wild-type hosts (20, 26, 73). E_{35}–GFP fusions have been constructed which are fully

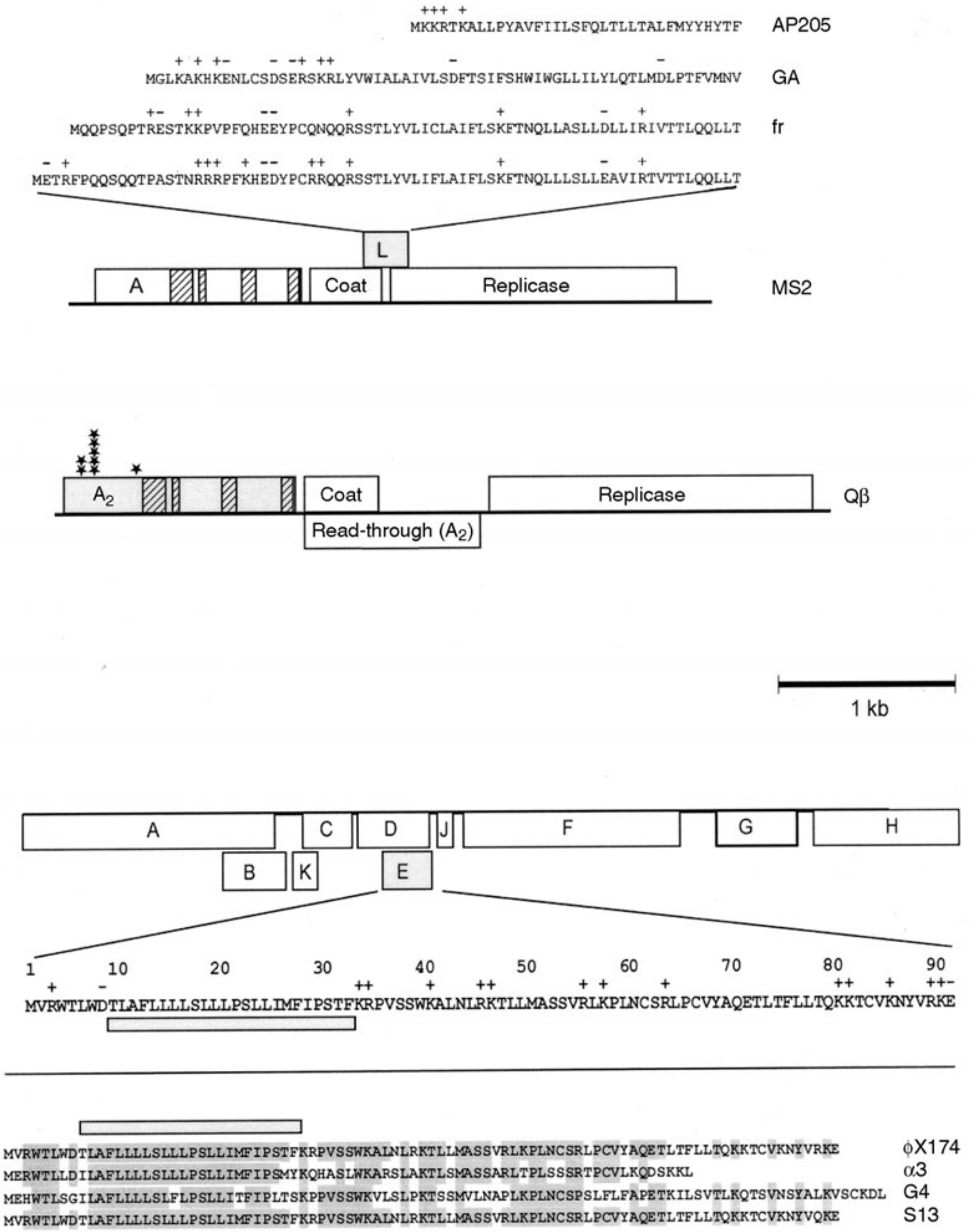

MKKRTKALLPYAVFIILSFQLTLLTALFMYYHYTF
AP205
MGLKAKHKENLCSDSERSKRLYVWIALAIVLSDFTSIFSHWIWGLLILYLQTLMDLPTFVMNV
GA
MQQPSQPTRESTKKPVPFQHEEYPCQNQQRSSTLYVLICLAIFLSKFTNQLLASLLDLLIRIVTTLQQLLT
fr
METRFPQQSQQTPASTNRRRPFKHEDYPCRRQQRSSTLYVLIFLAIFLSKFTNQLLLSLLEAVIRTVTTLQQLLT
L
A
Coat
Replicase
MS2
A₂
Coat
Replicase
Qβ
Read-through (A₂)
1 kb
A
C
D
J
F
G
H
B
K
E
1
10
20
30
40
50
60
70
80
90
MVRWTLWDTLAFLLLLSLLLPSLLIMFIPSTFKRPVSSWKALNLRKTLLMASSVRLKPLNCSRLPCVYAQETLTFLLTQKKTCVKNYVRKE
MVRWTLWDTLAFLLLLSLLLPSLLIMFIPSTFKRPVSSWKALNLRKTLLMASSVRLKPLNCSRLPCVYAQETLTFLLTQKKTCVKNYVRKE
ϕX174
MERWTLLDILAFLLLLSLLLPSLLIMFIPSMYKQHASLWKARSLAKTLSMASSARLTPLSSSRTPCVLKQDSKKL
α3
MEHWTLSGILAFLLLLSLFLPSLLITFIPLTSKPPVSSWKVLSLPKTSSMVLNAPLKPLNCSPSLFLFAPETKILSVTLKQTSVNSYALKVSCKDL
G4
MVRWTLWDTLAFLLLLSLLLPSLLIMFIPSTFKRPVSSWKALNLRKTLLMASSVRLKPLNCSRLPCVYAQETLTFLLTQKKTCVKNYVQKE
S13

Figure 10-9 Single-gene lysis systems in ssDNA and ssRNA phages. Lysis genes are shaded. Top: Map of MS2, the prototype group I ssRNA phage, with the L sequence displayed above the map. L sequences from the group I phage fr (3), the group II phage GA (61), and the *Acinetobacter phage* AP205 (69) are arrayed above MS2 L for comparison. Middle: Map of Qβ, the prototype group III ssRNA phage (105). The regions of close sequence similarity between Qβ A_2 and MS2 A are hatched, and the locations of the *por* mutations, selected for plaque-formation on a *rat* host, are indicated by stars (14) (I.-N. Wang and R. Young, unpublished data). Bottom: Map of the circular ssDNA phage ϕX174, shown linearized at the start of the *A* gene. The sequence of the E lysis protein is displayed below the map, along with an alignment of the ortholog E proteins from the related phages (top to bottom) α3, G4 and S13 (47, 121).

active lytically; fluorescence microscopy shows clearly that the hybrid E protein is uniformly distributed in the membrane (W. D. Roof and R. Young, unpublished data).

Early attempts at defining the target of E by isolating resistant host mutants were frustrated by the fact that knockout mutations in a nonessential host gene, *slyD*, abolished E-mediated lysis (73). SlyD is a peptidyl-prolyl cis-transferase isomerase (PPIase), orthologous to mammalian FKBP, the receptor for the immunosuppressant drugs FK506 and rapamycin (93). In the absence of SlyD, the E protein is very unstable, with a half-life of about 2 minutes (11). ϕX174 successfully infects *slyD* cells but, instead of

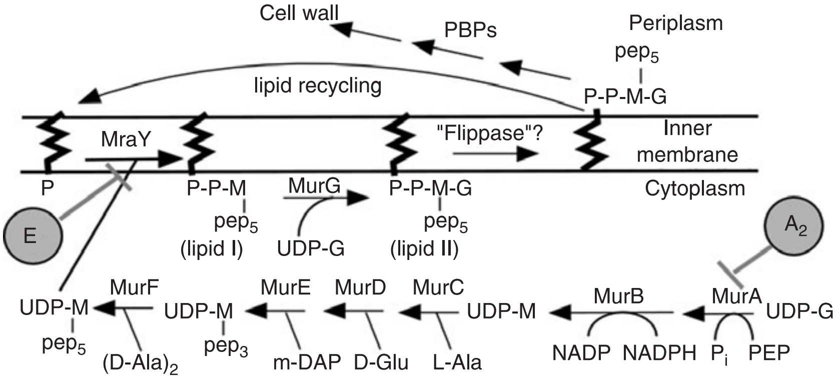


Figure 10-10 Phage amurins inhibit different steps in cell wall synthesis. A simplified version of the cell wall synthesis pathway, adapted from Nanninga (78). The jagged line in the membrane represents undecaprenylphosphate. Abbreviations used are: G, GlcNAc; M, MurNAc; pep_3, tripeptide; pep_5, pentapeptide; PBPs, penicillin-binding proteins. Two different steps are specifically blocked by the ϕXl74 and Qβ amurins (see text).

lysing the host cell, accumulates cytoplasmically to very high levels (94). ϕX174 *pos* mutants (*p*lates *o*n *s*lyD) were isolated which overcome the lysis block; the *pos* mutations are *E* mutations which increase the synthesis rate by about 10-fold, although the instability in the absence of SlyD is retained (11). An *Epos* allele cloned in a plasmid expression vector was used to select lysis-resistant mutants which mapped to *mraY*, a conserved membrane-embedded enzyme which catalyzes the formation of the first lipid-linked step in peptidoglycan biosynthesis (figure 10-10). The mutations map to TMD5 and TMD9 of MraY and block the specific inhibition mediated in vivo and in vitro by E, which does not inhibit a related membrane-embedded enzyme, Rfe, that also uses undecaprenol phosphate as a recipient and a UDP-activated donor molecule (10, 12). The basis of the enzyme inhibition is unknown, but it is worth noting that the C_{55} carrier is in very limited quantities in the membrane, so it is possible that the hydrophobic membrane domain of E competes for the carrier binding site. In any case, E is the first of a new class of lysis proteins, the amurins, which cause lysis by inhibiting a specific enzymatic step in murein synthesis.

In terms of how E-mediated lysis might be regulated, there is no evidence for regulation of *E* expression in ϕX174, where all transcripts include *E* (39). Lysis by E, as with the fungal antibiotic mureidomycin, a specific MraY-inhibitor, requires cell division and appears to be effected by a failed septation event (24, 64). This is consistent with the finding that *fts* mutants at the nonpermissive temperature are resistant to E-mediated lysis (115). Thus it would appear that the infective cycle of ϕX174 is defined by the cell-cycle stage at which the infection begins. Unless there are unknown compensatory strategies, phages which infect early after septation should have a longer vegetative phase and higher burst than phages which infect late in infection. It has not been ruled out that SlyD may mediate interactions with host physiology and thus provide a conduit for information relevant to the timing of lysis; interestingly, overproduction of SlyD leads to an immediate and reversible block in septation (92).

Qβ A_2

Like its orthologue A of the *Leviviridae*, the A_2 protein of Qβ and other *Alloleviviridae* is a >40 kDa single-copy component of the virion, with multiple roles including binding to the male pilus, RNA-binding during capsid assembly, protection of the RNA 3′ terminus, and penetration, with the RNA, into the cytoplasm of the host (105). Unlike MS2 A, however, A_2 is also the sole lysis protein of the phage (114). The discovery that ϕX174 E was an amurin, and evidence associating A_2 with a blockage of cell wall synthesis, suggested that A_2 might be another inhibitor of the murein synthesis pathway (13, 14, 82). Unlike E, however, no soluble lipid-linked or soluble precursors accumulated during A_2-mediated lysis, suggesting that an early step in the pathway was blocked. *rat* mutants (*r*esistant to *A*-*t*wo) were isolated as survivors of induction of a cloned A_2 gene (13). Eight independent, spontaneous *rat* mutants have been obtained and all are identical, with a Leu138Gln change in MurA, the conserved enzyme catalyzing the committed step of murein biosynthesis (14). Efforts to demonstrate MurA inhibition with purified A_2 have been frustrated by the latter's extreme insolubility even in 8 M urea, but, remarkably, high-titer suspensions of virions, each bearing a single copy of A_2, cause inhibition of MurA, but not $MurA^{rat1}$. The ability of the virion-mounted A_2 to inhibit MurA despite supersaturating concentrations of both substrates suggests the inhibition is noncompetitive or uncompetitive (14). The position of the *rat* mutation corresponds to an outside surface of the enzyme, far from the active site cleft.

Attempts using in vitro mutagenesis to separate the lysis function from the other functions of A_2 were unsuccessful (114). Comparison of the MS2 A and Qβ A_2 sequences are not informative as to a likely localization of a lysis domain because the two sequences are so dramatically divergent. However, Qβ*por* (*p*lates *o*n *r*at) mutants map to the N-terminal 120 residues of A_2, which is also one of the regions lacking similarity to A, suggesting that this portion of the protein constitutes the amurin domain (figure 10-9) (14).

The presence of the MurA inhibitor domain on the surface of the virion provides an obvious lysis control mechanism; when virions accumulate to sufficient levels to block the available MurA, cell wall synthesis stops and lysis ensues, if the cells are in growth phase. This link between virion content and lysis is broken, however, if cells are no longer dividing; interestingly, grossly higher virion yields can be obtained if late logarithmic cultures are infected and incubated until saturation (105).

MS2 L

The MS2 *L* gene is embedded across the end of the *coat* and beginning of the *rep* genes of the *Leviviridae* (figure 10-9). The local RNA secondary structure and the dynamics of ribosome movement in the immediate region of the start codon are critical for proper expression of the lysis gene, which has its translational initiation region occluded by a large secondary structure. The importance of this structure for the biological fitness of the phage, presumably because it is required to avoid inappropriately high synthesis of the L protein and thus the premature onset of lysis, has been shown in elegant suppressor studies by van Duin's group, who eliminated the secondary structure by a series of wobble-position changes within the *coat* gene (70). Serial passage of the mutated phage resulted in the accumulation of multiple suppressor mutations, again in wobble positions, that restored the secondary structure repressing *L* translation .

The predicted L product is virtually a mirror-image of E, in that its 75 residues are primarily hydrophilic and cationic in the first two thirds of its predicted sequence, and mostly hydrophobic in the last third (figure 10-9). Deletion analysis suggests that its lytic capacity resides in the 32 C-terminal residues (9). Unlike E and A_2, induction of a cloned *L* gene does not lead to an inhibition of incorporation of ^{3}H-diaminopimelate into murein before lysis (T. G. Bernhardt and R. Young, unpublished data). Thus, if L is also an amurin, it must affect a step beyond the first covalent attachment of the precursor disaccharide–pentapeptide into the murein layer. The L protein has been reported to be localized to zones of adhesion, to require the host membrane-derived oligosaccharide system for its lytic function, and to induce the host autolytic system (59, 106, 107). A synthetic polypeptide corresponding to its C-terminal 25 residues has been shown to permeabilize liposomes to fluorescent dyes (48). Other than the deletion analysis described above, no mutational study of L has been reported, nor has there been a genetic analysis to find host mutants with altered L-sensitivity. In the absence of such studies, it is unclear whether these diverse phenomena are directly related to the lytic function of L.

Perspectives and Directions

In recent years remarkable strides have been made in our understanding of the molecular basis of phage lysis and its regulation. The progress has raised many new questions. That there are at least two general strategies for lysis is clear. All complex phages seem to use holin–endolysin lysis, whereas two of the prototype small ssDNA and ssRNA phages encode proteins which act as inhibitors of cell wall synthesis. The diversity of holins has always been a stunning, and somewhat daunting feature of lysis, suggesting that there may be several fundamentally different lysis mechanisms. Nevertheless, the basic features of all these systems are still comfortingly common, especially the ability to be triggered by energy poisons. It will be useful to assess whether class II and class III holins share the "kill without warning" property of the S holin, as they should if our understanding of what drives holin evolution is correct. In addition, although the discovery of the secretory endolysins means that the muralytic enzymes can no longer be regarded as "dumb" reporter functions for the activity of holins, it has perhaps further brought into focus how important the timing function is for the lytic event that terminates all dsDNA phage infections. Holins apparently have evolved, perhaps independently and at multiple times, to provide a temporal schedule for lysis and to allow the optimization of that schedule. Competition experiments to test the fitness of various holin genes and series of timing-mutant alleles of a particular holin should be instructive. In terms of the mechanism of holin function, it is clear that the next step must include cytological localization of the holins to assess whether patches or rafts are formed in the membrane during the vegetative cycle. In vitro experiments in which holin function is reconstituted in artificial lipid vesicles should help clarify the roles of concentration, oligomerization, and membrane energization in the hole-formation process, as well as providing structural insight into the nature of the lesions. These same systems can also be exploited for investigating the mechanisms by which antiholins block holin function. The diversity of topologies available to the known antiholins suggest that a wealth of specific interactions underlie the regulatory properties of these molecules, interactions which should be fertile areas for genetic analysis.

With regard to the single-gene systems, it will be interesting to see whether MS2 L is also an amurin, or if a third general strategy is available, perhaps a "magic button" that allows the induction of so-called autolysis without disturbance of murein synthesis. It should be noted that a wide variety of ssRNA phages have been isolated against a number of different R factors (23). It is a reasonable expectation that some of these ssRNA phages may have evolved amurins against targets other than MurA. Recently, MH2K, a lytic *Microvirus* of *Bdellovibrio*, was isolated and sequenced. MH2K lacks a scaffolding protein gene equivalent to *D* (25) and, consequently, lacks the *E* lysis cistron embedded in *D* in ϕX174 and its coliphage relatives. Instead, the candidate lysis genes are short open reading frames embedded in other essential MH2K genes. It will be interesting to test whether this independently evolving lysis gene also targets MraY. Interestingly, a *Microvirus* sequence for the wall-less intracellular bacterium *Chlamydia* has no obvious reading frame available for a lysis gene (103). It is unknown how a phage can cause lysis of a host cell that grows, without a cell wall, in the iso-osmotic environment of a mammalian cell cytoplasm.

Acknowledgments

The authors wish to thank all the past and present members of the Young laboratory, as well as the growing international community of phage biologists and lysis *aficionados*, who have continually provided encouragement, stimulation, and frequently unpublished or pre-publication information. The work reported from the authors' laboratory was supported by NIH grant GM27099 to R. Y., an award from the Robert A. Welch Foundation, and funds from both Texas A&M University and the Texas Agricultural Experiment Station.

References

1. Abedon, S. T. 1989. Selection for bacteriophage latent period length by bacterial density: a theoretical examination. Microbiol. Ecol. 18:79–88.
2. Abedon, S. T. 1994. Lysis and the interaction between free phages and infected cells, pp. 397–405. *In* J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, K. Carlson, and E. S. Miller (eds.) Molecular Biology of Bacteriophage T4. American Society for Microbiology, Washington, D.C.
3. Adhin, M. R., A. Avots, V. Berzin, G. P. Overbeek, and J. van Duin. 1990. Complete nucleotide sequence of a group I RNA bacteriophage fr. Biochim. Biophys. Acta 1050: 104–109.
4. Altieri, M., J. L. Suit, M.-L. Fan, and S. E. Luria. 1986. Expression of the cloned ColE1 *kil* gene in normal and *kil*-R *Escherichia coli*. Proc. Natl. Acad. Sci. USA 168:648–654.
5. Atkins, J. F., J. A. Steitz, C. W. Anderson, and P. Model. 1979. Binding of mammalian ribosomes to MS2 phage RNA reveals an overlapping gene encoding a lysis function. Cell 18:247–256.
6. Barenboim, M., C.-Y. Chang, F. dib Hajj, and R. Young. 1999. Characterization of the dual start motif of a class II holin gene. Mol. Microbiol. 32:715–727.
7. Benjamin, R. C., W. P. Fitzmaurice, P. C. Huang, and J. J. Scocca. 1984. Nucleotide sequence of cloned DNA segments of the *Haemophilus influenzae* bacteriophage HP1c1. Gene 31:173–185.
8. Beremand, M. N., and T. Blumenthal. 1979. Overlapping genes in RNA phage: a new protein implicated in lysis. Cell 18:257–266.
9. Berkhout, B., M. H. de Smit, R. A. Spanjaard, T. Blom, and J. van Duin. 1985. The amino terminal half of the MS2-coded lysis protein is dispensable for function: implications for our understanding of coding region overlaps. EMBO J. 4:3315–3320.
10. Bernhardt, T. G., W. D. Roof, and R. Young. 2000. Genetic evidence that the bacteriophage ϕX174 lysis protein inhibits cell wall synthesis. Proc. Natl. Acad. Sci. USA 97: 4297–4302.
11. Bernhardt, T. G., W. D. Roof, and R. Young. 2002. The *Escherichia coli* FKBP-type PPIase SlyD is required for the stabilization of the E lysis protein of bacteriophage ϕX174. Mol. Microbiol. 45:99–108.
12. Bernhardt, T. G., D. K. Struck, and R. Young. 2001. The lysis protein E of ϕX174 is a specific inhibitor of the MraY-catalyzed step in peptidoglycan synthesis. J. Biol. Chem. 276:6093–6097.
13. Bernhardt, T. G., I. N. Wang, D. K. Struck, and R. Young. 2001. A protein antibiotic in the phage Qβ virion: diversity in lysis targets. Science 292:2326–2329.
14. Bernhardt, T. G., I.-N. Wang, D. K. Struck, and R. Young. 2002. Breaking free: "protein antibiotics" and phage lysis. Res. Microbiol. 153:493–501.
15. Bienkowska-Szewczyk, K., B. Lipinska, and A. Taylor. 1981. The *R* gene product of bacteriophage λ is the murein transglycosylase. Mol. Gen. Genet. 184:111–114.
16. Blasband, A. J., W. R. Marcotte, Jr., and C. A. Schnaitman. 1986. Structure of the *lc* and *nmpC* outer membrane porin protein genes of lambdoid bacteriophage. J. Biol. Chem. 261:12723–12732.
17. Bläsi, U., C.-Y. Chang, M. T. Zagotta, K. Nam, and R. Young. 1990. The lethal λ *S* gene encodes its own inhibitor. EMBO J. 9:981–989.
18. Bläsi, U., P. Fraisl, C.-Y. Chang, N. Zhang, and R. Young. 1999. The C-terminal sequence of the lambda holin constitutes a cytoplasmic regulatory domain. J. Bacteriol. 181: 2922–2929.
19. Bläsi, U., G. Halfmann, and W. Lubitz. 1984. Induction of autolysis of *Escherichia coli* by ϕX174 gene *E* product, pp. 213–218. *In* C. Nombela (ed.) Microbial Cell Wall Synthesis and Autolysis. Elsevier, New York.
20. Bläsi, U., and W. Lubitz. 1985. Influence of C-terminal modifications of ϕX174 lysis gene *E* on its lysis-inducing properties. J. Gen. Virol. 66:1209–1213.
21. Bläsi, U., K. Nam, D. Hartz, L. Gold, and R. Young. 1989. Dual translational initiation sites control function of the λ *S* gene. EMBO J. 8:3501–3510.
22. Bonovich, M. T., and R. Young. 1991. Dual start motif in two lambdoid *S* genes unrelated to λ *S*. J. Bacteriol. 173: 2897–2905.
23. Bradley, D. E. 1976. Adsorption of the R-specific bacteriophage PR4 to pili determined by a drug resistance plasmid of the W compatibility group. J. Gen. Microbiol. 95:181–185.

24. Bradley, D. E., C. A. Dewar, and D. Robertson. 1969. Structural changes in *Escherichia coli* infected with a ϕX174 type bacteriophage. J. Gen. Virol. 5:113–121.
25. Brentlinger, K. L., S. Hafenstein, C. R. Novak, B. A. Fane, R. Borgon, R. McKenna, and M. Agbandje-McKenna. 2002. *Microviridae*, a family divided: isolation, characterization, and genome sequence of ϕMH2K, a bacteriophage of the obligate intracellular parasitic bacterium *Bdellovibrio bacteriovorus*. J. Bacteriol. 184:1089–1094.
26. Buckley, K. J., and M. Hayashi. 1986. Lytic activity localized to membrane-spanning region of ϕX174 E protein. Mol. Gen. Genet. 204:120–125.
27. Calamita, H. G., and R. J. Doyle. 2002. Regulation of autolysins in teichuronic acid-containing *Bacillus subtilis* cells. Mol. Microbiol. 44:601–606.
28. Calamita, H. G., W. D. Ehringer, A. L. Koch, and R. J. Doyle. 2001. Evidence that the cell wall of *Bacillus subtilis* is protonated during respiration. Proc. Natl. Acad. Sci. USA 98:15260–15263.
29. Cao, G., and R. E. Dalbey. 1994. Translocation of N-terminal tails across the plasma membrane. EMBO J. 13:4662–4669.
30. Casjens, S. R., K. Eppler, R. Parr, and A. R. Poteete. 1989. Nucleotide sequence of the bacteriophage P22 gene *19* to *3* region: identification of a new gene required for lysis. Virology 171:588–598.
31. Chang, C.-Y. 1994. Synthesis, Function and Regulation of the Lambda Holin. Texas A&M University, Collge Station, Texas.
32. Chang, C.-Y., K. Nam, U. Bläsi, and R. Young. 1993. Synthesis of two bacteriophage λ S proteins in an *in vivo* system. Gene 133:9–16.
33. Chang, C.-Y., K. Nam, and R. Young. 1995. *S* gene expression and the timing of lysis by bacteriophage λ. J. Bacteriol. 177:3283–3294.
34. Cheng, X., X. Zhang, J. W. Pflugrath, and F. W. Studier. 1994. The structure of bacteriophage T7 lysozyme, a zinc amidase and an inhibitor of T7 RNA polymerase. Proc. Natl. Acad. Sci. USA 91:4034–4038.
35. Cohen, L. W., M. R. Showers, and W. D. Andrus. 1971. Delayed lysis in *Salmonella* phage P22: the continued division of mutant-infected cells actively producing phage. Virology 54:848–852.
36. Cohen, S. S. 1968. Virus-Induced Enzymes. Columbia University Press, New York.
37. Croux, C., and J. L. Garcia. 1991. Sequence of the *lyc* gene encoding the autolytic lysozyme of *Clostridium acetobutylicum* ATCC824: comparison with other lytic enzymes. Gene 104:25–31.
38. Deaton, J., J. Sun, A. Holzenburg, D. K. Struck, J. Berry, and R. Young. 2004.Functional bacteriorhodopsin is efficiently solubilized and delivered to membranes by the chaperonin GroEL. Proc. Natl. Acad. Sci. USA 101: 2281–2286.
39. Denhardt, D. T., D. Dressler, and D. S. Ray. 1978. The Single-Stranded DNA Phages. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
40. Diaz, E., R. Lopez, and J. L. Garcia. 1990. Chimeric phage-bacterial enzymes—a clue to the modular evolution of genes. Proc. Natl. Acad. Sci. USA 87:8125–8129.
41. Diaz, E., R. Lopez, and J. L. Garcia. 1991. Chimeric pneumococcal cell-wall lytic enzymes reveal important physiological and evolutionary traits. J. Biol. Chem. 266:5464–5471.
42. Dressman, H. K., and J. W. Drake. 1999. Lysis and lysis inhibition in bacteriophage T4: *rV* mutations reside in the holin *t* gene. J. Bacteriol. 181:4391–4396.
43. Evrard, C., J. Fastrez, and J. P. Declercq. 1998. Crystal structure of the lysozyme from bacteriophage lambda and its relationship with V and C-type lysozymes. J. Mol. Biol. 276:151–164.
44. Field, C. B., M. J. Behrenfeld, J. T. Randerson, and P. Falkowski. 1998. Primary production of the biosphere: integrating terrestrial and oceanic components. Science 281:237–240.
45. Fuhrman, J. A. 1999. Marine viruses and their biogeochemical and ecological effects. Nature 399:541–548.
46. Fung, D. C., and H. C. Berg. 1995. Powering the flagellar motor of *Escherichia coli* with an external voltage source. Nature 375:809–812.
47. Godson, G. N., J. C. Fiddes, B. G. Barrell, and F. Sanger. 1978. Comparative DNA sequence analysis of the G4 and ϕX174 genomes., pp. 51–86. *In* D. T. Denhardt, D. Dressler, and D. S. Ray (eds.) The Single-Stranded DNA Phages. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
48. Goessens, W. H. F., A. J. M. Driessen, J. Wilschut, and J. van Duin. 1988. A synthetic peptide corresponding to the C–terminal 25 residues of phage MS2-coded lysis protein dissipates the proton-motive force in *Escherichia coli* membrane vesicles by generating hydrophilic pores. EMBO J. 7:867–873.
49. Goldberg, A. R., and M. Howe. 1969. New mutations in the *S* cistron of bacteriophage lambda affecting host cell lysis. Virology 38:200–202.
50. Graschopf, A., and U. Bläsi. 1999. Functional assembly of the lambda S holin requires periplasmic localization of its N-terminus. Arch. Microbiol. 172:31–39.
51. Graschopf, A., and U. Bläsi. 1999. Molecular function of the dual-start motif in the λ S holin. Mol. Microbiol. 33: 569–582.
52. Gründling, A., U. Bläsi, and R. Young. 2000. Biochemical and genetic evidence for three transmembrane domains in the class I holin, λ S. J. Biol. Chem. 275:769–776.
53. Gründling, A., U. Bläsi, and R. Young. 2000. Genetic and biochemical analysis of dimer and oligomer interactions of the λ S holin. J. Bacteriol. 182:6082–6090.
54. Gründling, A., M. D. Manson, and R. Young. 2001. Holins kill without warning. Proc. Natl. Acad. Sci. USA 98: 9348–9352.
55. Gründling, A., D. L. Smith, U. Bläsi, and R. Young. 2000. Dimerization between the holin and holin inhibitor of phage lambda. J. Bacteriol. 182:6075–6081.
56. Hanych, B., S. Kedzierska, B. Walderich, B. Uznanski, and A. Taylor. 1993. Expression of the *Rz* gene and the overlapping *Rz1* reading frame present at the right end of the bacteriophage lambda genome. Gene 129:1–8.
57. Harris, A. W., D. W. A. Mount, C. R. Fuerst, and L. Siminovitch. 1967. Mutations in bacteriophage lambda affecting host cell lysis. Virology 32:553–569.

58. Hershey, A. D. 1946. Mutation of bacteriophage with respect to type of plaque. Genetics 31:620–640.
59. Höltje, J.-V., W. Fiedler, H. Rotering, B. Walderich, and J. van Duin. 1988. Lysis induction of *Escherichia coli* by the cloned lysis protein of the phage MS2 depends on the presence of osmoregulatory membrane-derived oligosaccharides. J. Biol. Chem. 263:3539–3541.
60. Hutchison, C. A. III, and R. L. Sinsheimer. 1966. The process of infection with bacteriophage ϕX174. X. Mutations in a ϕX lysis gene. J. Mol. Biol. 18:429–447.
61. Inokuchi, Y., R. Takahashi, T. Hirose, S. Inayama, A. B. Jacobson, and A. Hirashima. 1986. The complete nucleotide sequence of the group II RNA coliphage GA. J. Biochem. (Tokyo) 99:1169–1180.
62. Isenbarger, T. A., and M. P. Krebs. 1999. Role of helix–helix interactions in assembly of the bacteriorhodopsin lattice. Biochemistry 38:9023–9030.
63. Isenbarger, T. A., and M. P. Krebs. 2001. Thermodynamic stability of the bacteriorhodopsin lattice as measured by lipid dilution. Biochemistry 40:11923–11931.
64. Isono, F., M. Inukai, S. Takahashi, T. Haneishi, T. Kinoshita, and T. Kuwano. 1989. Mureidomycins A–D, novel peptidylnucleoside antibiotics with spheroplast forming activity. III. Biological properties. J. Antibiot. 42:674–679.
65. Johnson-Boaz, R., C.-Y. Chang, and R. Young. 1994. A dominant mutation in the bacteriophage lambda *S* gene causes premature lysis and an absolute defective plating phenotype. Mol. Microbiol. 13:495–504.
66. Joliffe, L. K., R. J. Doyle, and U. N. Streips. 1981. The energized membrane and cellular autolysis in *Bacillus subtilis*. Cell 25:753–763.
67. Josslin, R. 1970. The lysis mechanism of phage T4: mutants affecting lysis. Virology 40:719–726.
68. Kedzierska, S., A. Wawrzynow, and A. Taylor. 1996. The *Rz1* gene product of bacteriophage lambda is a lipoprotein localized in the outer membrane of *Escherichia coli*. Gene 168:1–8.
69. Klovins, J., G. P. Overbeek, S. H. van den Worm, H. W. Ackermann, and J. van Duin. 2002. Nucleotide sequence of a ssRNA phage from *Acinetobacter*: kinship to coliphages. J. Gen. Virol. 83:1523–1533.
70. Klovins, J., N. A. Tsareva, M. H. de Smit, V. Berzins, and J. van Duin. 1997. Rapid evolution of translational control mechanisms in RNA genomes. J. Mol. Biol. 265:372–384.
71. Lubitz, W., G. Halfmann, and R. Plapp. 1984. Lysis of *Escherichia coli* after infection with ϕX174 depends on the regulation of the cellular autolytic system. J. Gen. Microbiol. 130:1079–1087.
72. Lubitz, W., R. E. Harkness, and E. E. Ishiguro. 1984. Requirement for a functional host cell autolytic enzyme system for lysis of *Escherichia coli* by bacteriophage ϕX174. J. Bacteriol. 159:385–387.
73. Maratea, D., K. Young, and R. Young. 1985. Deletion and fusion analysis of the ϕX174 lysis gene *E*. Gene 40: 39–46.
74. Markov, D., G. Christie, B. Sauer, R. Calendar, T. Park, T. R. Young, and K. Severinov. P2 growth restriction on an *rpoC* mutant is suppressed by alleles of the *Rz1* homolog *lysC*. J. Bacteriol. 186:4628–4637.
75. Moffatt, B. A., and F. W. Studier. 1987. T7 lysozyme inhibits transcription by T7 RNA polymerase. Cell 49:221–227.
76. Nam, K. 1991. Translational Regulation of the *S* Gene of Bacteriophage Lambda. Texas A&M University, College Station, Texas.
77. Nam, K., U. Bläsi, M. T. Zagotta, and R. Young. 1990. Conservation of a dual-start motif in P22 lysis gene regulation. J. Bacteriol. 72:204–211.
78. Nanninga, N. 1998. Morphogenesis of *Escherichia coli*. Microbiol. Mol. Biol. Rev. 62:110–129.
79. Navarre, W. W., H. Ton-That, K. F. Faull, and O. Schneewind. 1999. Multiple enzymatic activities of the murein hydrolase from staphylococcal phage ϕ11. Identification of a D-alanyl-glycine endopeptidase activity. J. Biol. Chem. 274:15847–15856.
80. Nelson, D., L. Loomis, and V. A. Fischetti. 2001. Prevention and elimination of upper respiratory colonization of mice by group A streptococci by using a bacteriophage lytic enzyme. Proc. Natl. Acad. Sci. USA 98: 4107–4112.
81. Owen, J. E., D. W. Schultz, A. Taylor, and G. R. Smith. 1983. Nucleotide sequence of the lysozyme gene of bacteriophage T4. Analysis of mutations involving repeated sequences. J. Mol. Biol. 165:229–248.
82. Ozaki, K., and R. C. Valentine. 1973. Inhibition of bacterial cell wall mucopeptide synthesis: a new function of RNA bacteriophage Qβ. Biochim. Biophys. Acta 304: 707–714.
83. Paddison, P., S. T. Abedon, H. K. Dressman, K. Gailbreath, J. Tracy, E. Mosser, J. Neitzel, B. Guttman, and E. Kutter. 1998. The roles of the bacteriophage T4 *r* genes in lysis inhibition and fine-structure genetics: a new perspective. Genetics 148:1539–1550.
84. Palmer, M. 2001. The family of thiol-activated, cholesterol-binding cytolysins. Toxicon 39:1681–1689.
85. Parreira, R., C. Sao-Jose, A. Isidro, S. Domingues, G. Vieira, and M. A. Santos. 1999. Gene organization in a central DNA fragment of *Oenococcus oeni* bacteriophage fOg44 encoding lytic, integrative and non-essential functions. Gene 226:83–93.
86. Plunkett, G. III, D. J. Rose, T. J. Durfee, and F. R. Blattner. 1999. Sequence of Shiga toxin 2 phage 933W from *Escherichia coli* O157:H7: Shiga toxin as a phage late-gene product. J. Bacteriol. 181:1767–1778.
87. Poteete, A. R., and L. W. Hardy . 1994. Genetic analysis of bacteriophage T4 lysozyme structure and function. J. Bacteriol. 176:6783–6788.
88. Raab, R., G. Neal, C. Sohaskey, J. Smith, and R. Young. 1988. Dominance in lambda *S* mutations and evidence for translational control. J. Mol. Biol. 199:95–105.
89. Ramanculov, E. R., and R. Young. 2001. Functional analysis of the T4 *t* holin in a lambda context. Mol. Genet. Genomics 265:345–353.
90. Redfield, R. J., and A. M. Campbell. 1987. Structure of cryptic lambda prophages. J. Mol. Biol. 198:393–404.
91. Rietsch, A., and U. Blasi. 1993. Non-specific hole formation in the *Escherichia coli* inner membrane by lambda S proteins S independent of cellular *secY* and *secA* functions and of the proportion of membrane acidic phospholipids. FEMS Microbiol. Lett. 107:101–105.
92. Roof, W. D., H. Q. Fang, K. D. Young, J. Sun, and R. Young. 1997. Mutational analysis of *slyD*, an *Escherichia coli* gene encoding a protein of the FKBP immunophilin family. Mol. Microbiol. 25:1031–1046.

93. Roof, W. D., S. M. Horne, K. D. Young, and R. Young. 1994. *slyD*, a host gene required for ϕX174 lysis, is related to the FK506-binding protein family of peptidyl-prolyl *cis–trans*-isomerases. J. Biol. Chem. 269: 2902–2910.
94. Roof, W. D., and R. Young. 1995. ?X174 lysis requires *slyD*, a host gene which is related to the FKBP family of peptidyl-prolyl *cis-trans* isomerases. FEMS Microbiol. Rev. 17:213–216.
95. São-José, C., R. Parreira, G. Vieira, and M. A. Santos. 2000. The N-terminal region of the *Oenococcus oeni* bacteriophage fOg44 lysin behaves as a bona fide signal peptide in *Escherichia coli* and as a cis-inhibitory element, preventing lytic activity on Oenococcal cells. J. Bacteriol. 182:5823–5831.
96. Schmidt, C., M. Velleman, and W. Arber. 1996. Three functions of bacteriophage P1 involved in cell lysis. J. Bacteriol. 178:1099–1104.
97. Sekiya, K., R. Satoh, H. Danbara, and Y. Futaesaku. 1993. A ring-shaped structure with a crown formed by streptolysin O on the erythrocyte membrane. J. Bacteriol. 175:5953–5961.
98. Sheehan, M. M., J. L. Garcia, R. Lopez, and P. Garcia. 1997. The lytic enzyme of the pneumococcal phage Dp-1: a chimeric lysin of intergeneric origin. Mol. Microbiol. 25:717–725.
98a. Ramanculov, E. R. and R. Young. 2001. An ancient player unmasked: T4 *rI* encodes a *t*-specific antiholin. Mol. Microbiol. 41:575–583.
99. Smith, D. L., D. K. Struck, J. M. Scholtz, and R. Young. 1998. Purification and biochemical characterization of the lambda holin. J. Bacteriol. 180:2531–2540.
100. Smith, D. L., and R. Young. 1998. Oligohistidine tag mutagenesis of the lambda holin gene. J. Bacteriol. 180: 4199–4211.
101. Steiner, M., and U. Bläsi. 1993. Charged amino-terminal amino acids affect the lethal capacity of lambda lysis proteins S107 and S105. Mol. Microbiol. 8:525–533.
102. Steiner, M., W. Lubitz, and U. Bläsi. 1993. The missing link in phage lysis of gram-positive bacteria: gene *14* of *Bacillus subtilis* phage ϕ29 encodes the functional homolog of lambda S protein. J. Bacteriol. 175:1038–1042.
103. Storey, C. C., M. Lusher, and S. J. Richmond. 1989. Analysis of the complete nucleotide sequence of Chp1, a phage which infects avian *Chlamydia psittaci*. J. Gen. Virol. 70:3381–3390.
104. Taylor, A., S. Kedzierska, and A. Wawrzynów. 1996. Bacteriophage λ lysis gene product modified and inserted into *Escherichia coli* outer membrane: Rz1 lipoprotein. Microb. Drug Resistance 2:147–153.
105. van Duin, J. 1988. Single-stranded RNA bacteriophages, pp. 117–167. *In* R. Calendar (ed.) The Bacteriophages. Plenum Press, New York.
106. Walderich, B., and J.-V. Höltje. 1989. Specific localization of the lysis protein of bacteriophage MS2 in membrane adhesion sites of *Escherichia coli*. J. Bacteriol. 171: 3331–3336.
107. Walderich, B., A. Ursinus-Wöosner, J. van Duin, and J.-V. Höltje. 1988. Induction of the autolytic system of *Escherichia coli* by specific insertion of bacteriophage MS2 lysis protein into the bacterial cell envelope. J. Bacteriol. 170: 5027–5033.
108. Walker, J. T., and D. H. Walker, Jr. 1980. Mutations in coliphage P1 affecting host cell lysis. J. Virol. 35:519–530.
109. Wang, I.-N., J. F. Deaton, and R. Young. 2003. Sizing the holin lesion with an endolysin–βgalactosidase fusion. J. Bacteriol. 185:779–787.
110. Wang, I.-N., D. E. Dykhuizen, and L. B. Slobodkin. 1996. The evolution of phage lysis timing. Evol. Ecol. 10: 545–558.
111. Wang, I.-N., D. L. Smith, and R. Young. 2000. Holins: the protein clocks of bacteriophage infections. Annu. Rev. Microbiol. 54:799–825.
112. Watson, J. D., N. H. Hopkins, J. W. Roberts, J. A. Steitz, and A. M. Weiner. 1987. Molecular Biology of the Gene, p. 507. Benjamin/Cummings, Menlow Park, Calif.
113. Wilhelm, S. W., and C. A. Suttle. 1999. Viruses and nutrient cycles in the sea. BioScience 49:781–783.
114. Winter, R. B., and L. Gold. 1983. Overproduction of bacteriophage Qβ maturation (A_2) protein leads to cell lysis. Cell 33:877–885.
115. Witte, A., E. Brand, P. Mayrhofer, F. Narandja, and W. Lubitz. 1998. Mutations in cell division proteins FtsZ and FtsA inhibit ϕX174 protein-E-mediated lysis of *Escherichia coli*. Arch. Microbiol. 170:259–268.
116. Witte, A., G. Schrot, P. Schon, and W. Lubitz. 1997. Proline 21, a residue within the alpha helical domain of ϕX174 lysis protein E, is required for its function in *Escherichia coli*. Mol. Microbiol. 26:337–346.
117. Witte, A., G. Wanner, U. Bläsi, G. Halfmann, M. Szostak, and W. Lubitz. 1990. Endogenous transmembrane tunnel formation mediated by ϕX174 lysis protein E. J. Bacteriol. 172:4109–4114.
118. Witte, A., G. Wanner, M. Sulzner, and W. Lubitz. 1992. Dynamics of ϕX174 protein E-mediated lysis of *Escherichia coli*. Arch. Microbiol. 157:381–388.
118a. Xu, M., A. Arulandu, D. K. Struck, S. Swanson, J. C. Sacchettini, and R. Young. 2005. Disulfide isomerization after membrane release of its SAR domain activates P1 lysozyme. Science 307:113–117.
118b. Xu, M., D. K. Struck, J. Deaton, I. N. Wang, and R. Young. 2004. The signal-arrest-release (SAR) sequence mediates export and control of the phage P1 endolysin. Proc. Natl. Acad. Sci., USA 101:6415–6420.
119. Young, R. 1992. Bacteriophage lysis: mechanism and regulation. Microbiol. Rev. 56:430–481.
120. Young, R. 2002. Bacteriophage holins: deadly diversity. J. Mol. Microbiol. Biotechnol. 4:21–36.
121. Young, R., I.-N. Wang, and W. D. Roof. 2000. Phages will out: strategies of host cell lysis. Trends Microbiol. 8: 120–128.
122. Young, R., S. Way, J. Yin, and M. Syvanen. 1979. Transposition mutagenesis of bacteriophage lambda: a new gene affecting cell lysis. J. Mol. Biol. 132:307–322.
123. Zhang, N., and R. Young. 1999. Complementation and characterization of the nested *Rz* and *Rz1* reading frames in the genome of bacteriophage lambda. Mol. Gen. Genet. 262:659–667.
124. Ziermann, R., B. Bartlett, R. Calendar, and G. E. Christie. 1994. Functions involved in bacteriophage P2-induced host cell lysis and identification of a new tail gene. J. Bacteriol. 176:4974–4984.

This page intentionally left blank

PART III

CUBIC AND FILAMENTOUS PHAGES

This page intentionally left blank

11

ϕX174 et al., the *Microviridae*

BENTLEY A. FANE, KARIE L. BRENTLINGER, APRIL D. BURCH,
MIN CHEN, SUSAN HAFENSTEIN, ERICA MOORE,
CHRISTOPHER R. NOVAK, AND ASAKO UCHIYAMA

Sometime in the 1920s, a plaque was placed in a test tube labeled #174. It must have had a certain luster. Since then that phage has challenged genetic, biochemical, and biophysical dogmas. From the very beginning, it was unique. It was unusually tiny, readily passing through the smallest of ultrafilters. Sertic and Bulgakov (111) used its size to define a "race": race X (ten). Hence the name became ϕX174. The first electron micrographs revealed a small, isometric particle without an elaborate, and to some unseemly, tail structure. As the study of molecular biology progressed, the particle refused to behave. Its genome was single-stranded DNA (117), not a double stranded molecule. Although the data used to draw the first genetic maps were accurate, revealing the existence of the overlapping reading frames, the cistrons were arranged in an orderly and linear manner (12). However, when the genome was sequenced (1, 107) the raw data were proved correct. This complex genetic arrangement was so unexpected that many suspected an extraterrestrial origin. As the *New York Times* reported the theory (127), an advanced race engineered ϕX174 and disseminated it into the cosmos where it would "persist until the evolution of intelligent life and finally of investigators interested in the genetics of phage." Attempts were made to find the hidden message in the genome, a holy truth or grail, maybe directions to an inhabited planet; however, none was uncovered. Perhaps the only aspect of ϕX174 that lends itself to science fiction is its name.

Since the last edition of *The Bacteriophages*, the results of ϕX174 and *Microviridae* research have continued to challenge paradigms. The atomic structures of the ϕX174 virion and grotesquely beautiful procapsid have been solved (35, 36, 92, 93, 94). As the genome sequence challenged genetic dogmas, the *structures* of the external scaffolding protein have challenged fundamental ideas of protein folding and particle morphogenesis. Furthermore, the results of biophysical studies suggest a scaffolding-like function for the genome (56) during the final stages of morphogenesis. Unlike large dsDNA viral morphogenesis, these final stages involve a radial collapse of the procapsid to form the virion, as opposed to an expansion. The characterization of a new *Microviridae* subfamily (20, 88, 104, 126) has offered insights into the mechanisms driving icosahedral single-stranded DNA viral evolution, which appear to be fundamentally different from those driving the evolution of the double-stranded DNA viruses (66, 67). And finally, studies of phage-mediated lysis mechanisms demonstrate how the *Microviridae* lyse cells without using lysozymes, but by "antibiotic-like" proteins (15, 16, 106). Quite frankly, ϕX174 is a rogue; however, to those who work with the *enfant terrible*, its recalcitrant nature is also its charm.

This chapter will concentrate on research conducted since the last edition of this book. Therefore, topics such has host cell recognition, genome penetration, gene expression, and DNA replication, which have not been active areas research in the last decade, will only be briefly summarized. Readers wishing for a more comprehensive review of these topics should consult the chapter by Masaki Hayashi (62) in the previous edition of this book.

Host Cell Recognition and Penetration

ϕX174 attaches to host cells via a sugar molecule in lipopolysaccharides (LPS) of Gram-negative bacteria (47, 73, 76, 79, 80). Bacterial strains that do not produce LPS containing specific terminal glucose and galactose molecules are resistant to ϕX174 infection (47). The *E. coli* gene *rfab*, located in the *rfa* cluster at 81 minutes, most likely encodes the requisite LPS biosynthetic enzyme (109). In solving the atomic structure of the ϕX174 virion, the crystallized particles were purified in sucrose gradients (92–94). As a result of this technique, glucose molecules were visualized bound to the coat protein in a defined position. The genes and gene products of ϕX174 are given in table 11-1. The six residues that constitute this binding site are strongly conserved in the other coliphage *Microviridae* (52, 83). More sophisticated

Table 11-1 The ϕX174 Gene Products

Protein	Function
A	Stage II and stage III DNA replication
A*	An unessential protein for viral propagation. It may play a role in the inhibition of host cell DNA replication and superinfection exclusion
B	Internal scaffolding protein, required for procapsid morphogenesis and the assembly of early morphogenetic intermediates. Sixty copies present in the procapsid
C	Facilitates the switch from stage II to stage III DNA replication. Required for stage III DNA synthesis
D	External scaffolding protein, required for procapsid morphogenesis. Two hundred and forty copies present in the procapsid
E	Host cell lysis
F	Major coat protein. Sixty copies present in the virion and procapsid
G	Major spike protein. Sixty copies present in the virion and procapsid
H	DNA pilot protein need for DNA injection, also called the minor spike protein. Twelve copies in the procapsid and virion
J	DNA binding protein, needed for DNA packaging. Sixty copies present in the virion
K	An unessential protein for viral propagation. It may play a role optimizing burst sizes in various hosts

crystallographic studies have elucidated the molecular basis of Ca^{2+} dependence for host cell recognition (73). Upon soaking the ion into the crystal, the atoms bind to defined locations at the 3-fold axes of symmetry. This causes the amino acid side chains of the glucose binding site residues to assume an ordered conformation.

The interaction between the glucose binding site and the LPS most likely accounts for the reversible reaction observed in kinetic experiments (75), which is followed by an irreversible reaction (77) the molecular basis of which remains obscure. DNA ejection can be simulated in vitro by high ionic conditions (138). Viral DNA is extruded from 5-fold vertices. In addition, ϕX174 host range mutations confer amino acid changes in the 5-fold related spike proteins G and H (98, 118, 136). LPS binding by these proteins has also been reported (78, 82). Host range coat protein mutations do not map to the strongly conserved glucose-binding site, but to a series of amino acids which trace the G protein around the 5-fold axes of symmetry (31). These results argue for the existence of a second host cell receptor, perhaps required for penetration. Although the identity of this second factor is unknown, a gene required for ϕX174 sensitivity (*phxB*) has been defined and mapped to minute 17 of the *E. coli* chromosome (96). In the *E. coli* genome sequence (18) there are several genes which could encode this second receptor or be involved in its synthesis. The tail-less ϕX174 penetration process may not differ significantly from that of the large-tailed bacteriophage that "walk" along the surface of the cell, via tailspike interactions, until they find a receptor for penetration. Instead of walking, ϕX174 may simply roll.

Electron micrographs of ϕX174 infections show the vast majority of adsorbed particles are embedded at points of adhesion between the cell wall and inner membrane (11, 21), suggesting the location of the hypothesized second receptor and indicating that the genome may be ejected directly into the cytoplasm. Penetration requires the viral H protein, also called the DNA pilot protein, which also enters into the cell (79, 80). In the atomic structure of the ϕX174 virion (73, 93, 94) diffuse density is located at each 5-fold vertex and has been attributed to protein H. A genetic analysis of a cold-sensitive H protein defective in DNA ejection supports this hypothesis (J. Oberste and B. A. Fane, unpublished data). Second-site suppressors of this allele map to two places in the major spike protein. One set of suppressing residues lines the channel that passes through each 5-fold vertex. The affected amino acids participate in 5-fold related contacts, hence maintaining the channel's integrity. The second set alters the interface between β-strands B and I of the G protein β-barrel, suggesting that conformational switches within this interface may mediate the opening of the channel.

DNA Replication

Positive polarity single-stranded DNA replication strategies are complex, occurring in three separate stages. These processes will be briefly summarized; a more complete discussion can be found in the previous edition of this book (62). Stage I DNA replication involves the conversion of the single-stranded genome into a covalently closed, double-stranded, circular molecule, called RF I DNA (replicative form one DNA). Purified single-stranded *Microviridae* DNA produces progeny in transfection experiments. Therefore, host cell proteins are both necessary and sufficient for stage I replication in vivo. The entire reaction has been reconstituted in vitro and requires 13 host proteins (84, 85, 114). A stem-loop structure in the FG intercistronic region serves as the n′ protein recognition site (113), and initiates the assembly of the primosome. As with most DNA priming structures, the stem-loop is not destabilized by single-stranded DNA binding protein (ssb). Additional proteins add to the complex in successive steps (6, 7, 89). After primosome assembly, the complex migrates along the single-stranded DNA in a 5′→3′ direction (5) synthesizing RNA primers for DNA replication. Addition of the holoenzyme leads to chain elongation.

During stage II DNA synthesis, RF I DNA is amplified. In addition to the 13 host cell proteins required for stage I replication, stage II replication is dependent on the viral A protein and the host cell helicase, *rep* protein (41). Replication of the (+) strand proceeds through a rolling circle mechanism (39). The viral A protein binds to the origin of replication, a 30-nucleotide sequence between residues 4299 and 4328, in the RF II DNA (10) and nicks it to initiate (+) strand synthesis (41, 135). The origin is separated into three functional domains. a binding domain, a cleavage recognition site, and an AT spacer region. The AT spacer region, which is located between the other two domains, is believed to facilitate helix unwinding. After nicking, protein A forms a covalent ester bond with the DNA (40). The host cell *rep* protein unwinds the helix and strand separation is stabilized by ssb (110). After one round of rolling circle synthesis, it cuts the newly generated origin and acts as a ligase, generating a covalently closed circular molecule (22, 39). Minus strand synthesis is mechanistically similar to stage I DNA synthesis.

Stage III DNA synthesis involves the concurrent synthesis and packaging of the single-stranded DNA genome. Viral procapsids and protein C are required for this reaction (61, 95). The viral C protein associates with proteins A and rep on the RF II DNA. The C protein may also serve as an inhibitor of double-stranded DNA synthesis (4). Aoyama and Hayashi (4) proposed a model in which ssb and protein C compete to bind single stranded DNA located in the protein A–rep–RF II complex. If ssb binds first, another round of stage II DNA synthesis occurs. Conversely, if protein C binds, stage III DNA and packaging ensues. This competition, however, can only occur at the initiation or reinitiation of DNA synthesis. Protein C will not inhibit stage II DNA synthesis after it has begun.

The protein A–C–rep–RF II DNA complex must dock to a viral procapsid for stage III DNA synthesis to commence. Procapsid binding does not occur in the absence of protein C (4). Second-site suppressors of mutant *rep* proteins, which abrogate procapsid binding, map to the A and viral coat proteins (43, 128, 133). The coat protein suppressors are clustered along the 2-fold axis of symmetry in the atomic structure of the virion (35, 36, 92–94), suggesting a location for the docking site. In addition, procapsids containing mutant external scaffolding proteins, which are also located at the 2-fold axes of symmetry, cannot be filled (see below). Again, packaging can be restored with mutations in protein A. Mechanistically stage III DNA synthesis is similar to the stage II (+) strand synthesis. Protein A nicks the origin (134) and forms a covalent ester bond with the DNA (40). After one round of rolling circle synthesis, it cuts the newly generated origin and acts as a ligase, generating a covalently closed circular molecule (22, 39). The packaging reaction is now complete. Even if the packaged genome has been altered to be less than unit length, reinitiation, as seen in headful packaging systems (87), does not occur (3). The origin is both necessary and sufficient for packaging specificity (3, 49). Any circular DNA molecule with *Microviridae* origin of replication can serve as a template (3, 56). The participation of the DNA-binding protein in packaging is discussed below.

Gene Expression

Since *Microviridae* contain single-stranded genomes of positive polarity, the negative strand must be synthesized before any genes can be transcribed and proteins translated. Unlike larger double-stranded DNA bacteriophages, *Microviridae* do not utilize trans-acting mechanisms to ensure temporal gene expression. Therefore, the timing and relative production of viral proteins is entirely dependent on cis-acting regulation signals: promoters, transcription terminators, mRNA stability sequences, and ribosome binding sites. Promoters are found upstream of genes A, B, and D (8, 9, 107, 119, 120–122), and terminators are found after genes J, F, G, and H (figure 11-1). The terminators are not 100% efficient, which leads to a wide variety of transcripts of varying lengths. However, there is a rough correlation between the abundance of a gene transcript and the amount of the

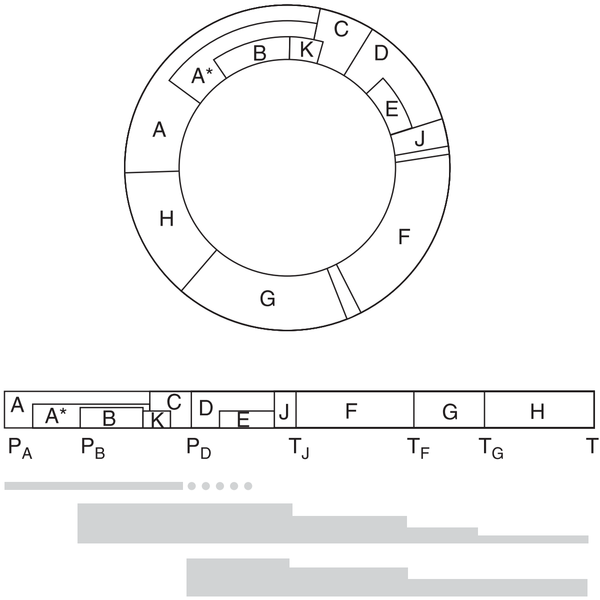


Figure 11-1 The genetic map of ϕX174. The promoters and transcription terminators are indicated on the linear map of ϕX174. Line thickness indicates the relative abundance of the transcripts. The gene A transcript is very unstable; the terminator for this transcript is unknown. Adapted from Hayashi et al. (62).

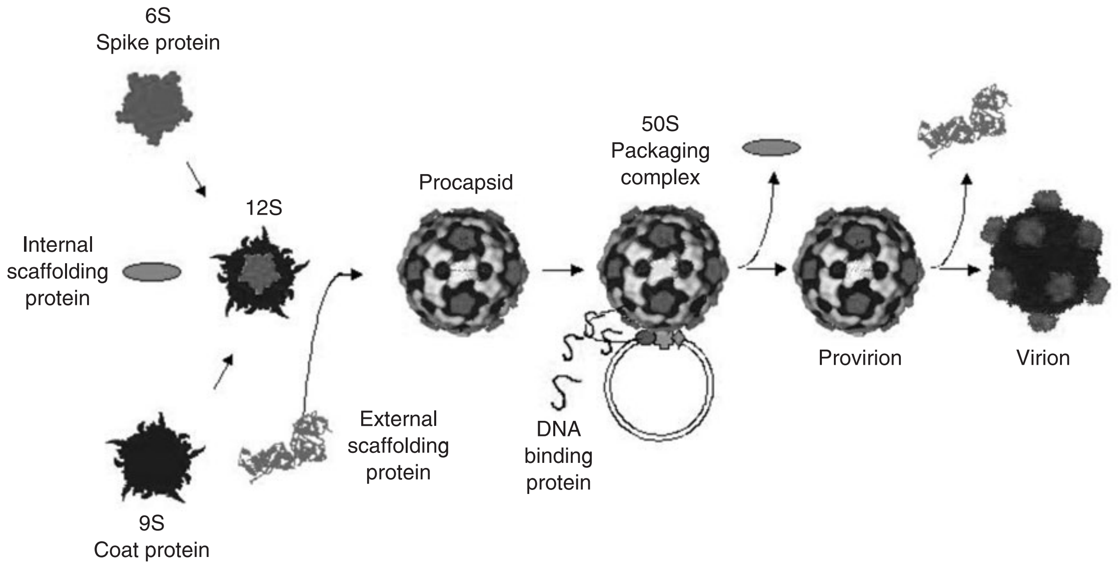


Figure 11-2 Phage β X174 morphogenesis.

encoded protein required for the viral life cycle (63). For example, gene D transcripts are the most abundant in the cell and protein D is the most abundant protein in the procapsid. Similarly, there are more gene F, J, and G transcripts than transcripts of gene H. The relative stoichiometry of these structural proteins is 5:5:5:1, respectively. Protein expression is also affected by the stability of the various mRNAs. Stability appears to be a function of 3′ end of the message (65) and each mRNA species decays with a characteristic rate (64). Transcripts of gene A decay very quickly (65), which may ensure that this nonstructural protein is not overexpressed. And finally, regulation can also be achieved at the translational level. Despite gene E's location within gene D, the most abundant transcript, few E proteins, which mediate cell lysis, are translated due to a weak ribosome binding site (17).

The Morphogenesis and Structures of the Virion and Procapsid

During the last decade, the atomic structure of several *Microviridae* virions and the ϕX174 procapsid have been solved (35, 36, 92–94). This, combined with well-established genetics and biochemistry, has made the ϕX174 system one of the most powerful for studying morphogenesis. While crystal structures provide a wealth of information, allowing the results of genetic experiments to be interpreted within a structural context, data are limited to the particle crystallized. Transient or less stable interactions between proteins are not always apparent (26, 35, 36, 72), but can be elucidated by the results of genetic and biochemical studies.

Pre-scaffolding Stages

As illustrated in figure 11-2, the first ϕX174 morphogenetic intermediates are the 9S and 6S particles, respective pentamers of the viral coat and spike proteins. These particles most likely self-assemble without the aid of scaffolding or host cell proteins such as groEL and ES (51, 68, 124). Chaperone-independence distinguishes the ϕX174 coat protein from those of larger bacteriophage (34, 53, 59, 97). The *groE* genes were first defined as host cell mutants that failed to support the growth of several double-stranded DNA bacteriophages (124). Although extensive searches have been conducted with ϕX174, mutations in molecular chaperones have never been recovered. However, mutations in other genes, such as the *rep* helicase, are abundant (128; M. Hayashi, personal communication).

The Internal Scaffolding Protein

In cells infected with nonsense or temperature-sensitive alleles of the internal scaffolding protein, protein B, 9S and 6S particles accumulate (42, 115, 132), demonstrating that pentamer formation is not a function of the first scaffolding-mediated steps of morphogenesis. Furthermore, the atomic structures of the *Microviridae* capsids reveal extensive 5-fold related contacts, suggesting a self-assembly mechanism (35, 36, 92–94). Pentamers formed in the absence of functional scaffolding proteins do not differ biologically from pentamers formed in its presence. Upon shifts to permissive temperatures in *tsB*- and *csB*-infected cells, 9S and 6S particles are efficiently chased into virions (42, 115).

After 9S pentamer formation, the internal scaffolding protein binds to the pentamer's underside and induces a conformational change in the particle. This change inhibits premature aggregation (115), and produces an assembly-competent state. B protein binding is both necessary and sufficient to allow interactions with the external scaffolding and major spike proteins (42, 132). Elucidating the structural changes of the conformational switch has proved difficult. Ideally one would compare the atomic structures of a naïve pentamer with a pentamer in the procapsid. However, assembly-naïve 9S particles aggregate in vitro, complicating crystal formation (R. McKenna, personal communication). Nevertheless, the results of second-site genetic analyses of a cold-sensitive B allele have offered some insights into the nature of the conformational switch (42, 45) and indicate that the morphogenetic changes occur on the outer surface of the coat protein.

Although morphogenesis does not continue past the first B-protein-mediated reaction in cells infected with this cold-sensitive B mutant, two lines of evidence suggest that the mutant protein retains some level of function, indicating a defect in conformational switching. First, 9S particles do not aggregate in vivo, suggesting that the *csB* protein can still inhibit the aggregation of coat protein pentamers. Second, the mutant is rescued by substitutions located on the outer surface of the coat protein, not the scaffolding–coat protein interface. The mutations are located within three distinct sequences of considerable homology, all found in loop regions of the protein, as opposed to the β-barrel core. These sequences may play a key role, perhaps as hinges, in mediating pentamer conformational switches. Coat protein mutations affecting other stages of morphogenesis—external scaffolding protein interactions, packaging complex recognition, B protein specificity, and provirion to virion transition—have also been isolated (28, 43, 46, 56, 81). All these mutations are found within the loop regions of the atomic structure, suggesting that, once folded, the contribution of the β-barrel core to morphogenesis is minimal.

Genetic analyses of the internal scaffolding protein have been impeded by a dearth of scaffolding protein missense mutations that confer morphogenetic defects. The dearth of such mutations in phage systems is rare. Although mutations in the gene B reading frame can also produce mutations in the gene A—proteins A* and K are nonessential (30, 129)—the results of genetic analyses indicate that genome organization is not responsible for this phenomenon. The B gene was cloned and mutagenized in a plasmid, thus removing any selective pressures that may be enforced by the overlapping reading frames. After mutagenesis, and reintroduction of the mutagenized plasmids into host cells, clones were screened for loss of the ability to complement *amB* phage. Although many noncomplementing B genes were recovered, all the mutations were nonsense or frameshift mutations. Furthermore, the mutagenized clones were screened for the ability to inhibit wild-type plaque formation, an assay for dominant lethal gene products. None were observed. The results of these experiments suggest that the lack of morphogenetic missense mutations is the result of a highly tolerant protein structure.

To test this hypothesis, the internal scaffolding proteins of the related bacteriophages α3 and G4 were cloned, expressed in vivo and assayed for the ability to cross-complement phages φX174, G4, and α3 *nullB* mutants. Despite the low sequence homology (30%) the proteins were, with one exception (see below), capable of cross-complementation, yielding not only procapsids but mature virions (28). However, in all cases morphogenesis was more efficient when directed by the indigenous protein. In essence, the “foreign” scaffolding proteins can be regarded as a “multiple mutant.” If a scaffolding protein in which 70% of the amino acids are altered is functional, the difficulty in obtaining defective proteins with single amino acid substitutions becomes apparent. These results also indicate that scaffolding proteins are inherently flexible. Experiments conducted with the analogous bovine herpesvirus 1 and herpes simplex virus (HSV) proteins yielded similar results (54), suggesting that flexibility may be a general property of internal scaffolding proteins. Considering the dynamics of viral assembly, some inherent flexibility is probably required. Internal scaffolding proteins must first assume a structure that directs the assembly of coat proteins into a rigid capsid. Afterwards, these proteins must assume an alternative structure, compact enough to be extruded through 2–3 nm pores, as observed in the φX174 and P22 systems (72, 102).

Evidence for inherent flexibility is also revealed in the φX174 procapsid structure. In the procapsid, the internal scaffolding protein binds to a cleft formed between α-helix 2 and the β-barrel of the coat protein (35, 36, 92). Coat protein binding is mediated by the last 24 amino acids of the scaffolding protein, which is the only portion of the proteins used in the cross-complementation studies that exhibit a high degree of sequence conservation. These interactions are primarily aromatic and comprise the most intimate B-coat protein contacts. The C-terminus is also the most ordered part of the protein. The first 60 amino acids of the protein yields primarily diffuse density, suggesting that interactions are variable and/or nonspecific in nature. In addition, the N-termini of the cross-functional scaffolding proteins are highly divergent. Therefore, it is unlikely that coat/N-terminal scaffolding interactions are governed by specific side chains. Again, structural studies with P22 and herpesvirus procapsids have yielded similar results. In cryo-image reconstructions, the C-termini of the proteins appear to produce ordered density, where the protein is in close contact with the capsid protein (99, 131, 141).

Data from three diverse systems—φX174, P22, and HSV-1—suggest that the C-termini of internal scaffolding proteins play a critical role in coat protein recognition. The importance of this region in the φX174 system is reinforced by both genetic and biochemical data. As stated above, the scaffolding proteins of the related phages G4, φX174, and α3 are able to cross-complement (25, 28). However, there was one instance in which cross-complementation was not observed. The ϕX174 protein cannot participate in the formation of the G4 procapsid. Characterization of the G4 morphogenetic pathway under these conditions revealed an accumulation of major coat and spike protein pentamers, indicating that morphogenesis terminated before the first internal scaffolding protein mediated reaction. In addition, the ϕX174 B protein does not inhibit wild-type G4 morphogenesis. These data suggest that the ϕX174 B and G4 coat proteins cannot interact. However, two G4 mutants (*ϕXB-utilizers*) that productively utilize the ϕX174 B protein were isolated. These mutations confer substitutions in the G4 coat protein that contact the C-terminal half of the B protein. In both instances, the substitutions create local coat protein sequences that are more ϕX174-like. One of these substitutions is located directly within the aromatic B protein-binding cleft. It confers a Ser → Phe substitution and most likely reflects the importance of aromatic interactions in coat recognition. The importance of C-termini interactions was further investigated by constructing chimeric B genes (25). The ϕXG4 B protein complements G4 *nullB* mutants, demonstrating that the inability of the ϕX174 B protein to interact with the G4 coat protein is a function of the C-terminus. In addition, when the C-terminus of any chimeric scaffolding protein was of the same origin as the viral coat protein, complementation was the most efficient.

For the most part, the function(s) of the N-termini remains obscure. The procapsid atomic structure suggests that the first 10 amino acids form an α-helix that self-associates across the 2-fold axis of symmetry. However, proteins lacking this α-helix function like the wild-type (25). This deletion may be tolerated because the external scaffolding protein also stabilizes the 2-fold axes of symmetry (35, 36). Cryo-image reconstructions of the α3 procapsid suggest that a large portion of the internal scaffolding protein, approximately residues 15–60, may be located at the 5-fold axes of symmetry, perhaps associated with the DNA pilot protein (14). A series of larger N-terminal deletion proteins have been constructed (100). The elimination of the first 38 amino acids supports large particle formation but terminates morphogenesis before virion production. The defective particles contain a full complement of coat and spike proteins but a drastically reduced amount of the H DNA pilot protein. These results support the hypothesis that a portion of the B protein may be located at the 5-fold axes of symmetry and indicate that the internal scaffold protein may play a role in H protein incorporation. Oddly, B proteins lacking the first 53 amino acids support virion production at temperatures above 33°C. Below 33°C, the mutant protein cannot support procapsid formation. Mutants that can utilize this deletion protein at lower temperatures have been isolated. The mutations map to dimerization domains in the external scaffolding protein. Perhaps a stronger external lattice can compensate for internal scaffolding proteins which cannot fully induce the conformational changes in coat pentamers needed to form capsid curvature.

The exact identity of the assembly intermediate between the 9S and 6S particles and the procapsid remains somewhat obscure. Unlike most viruses ϕX174 morphogenesis also relies on an external scaffolding protein (protein D) in order to assemble a pentameric intermediate into the procapsid. However, prior interaction with the internal scaffolding protein is required for subsequent coat protein interactions with the external scaffolding and spike proteins (115). In the absence of a functional D protein, 12S particles accumulate (132). These particles appear to have 5-fold rotational symmetry (62). Three chemically distinct 12S particles have been isolated (42, 45, 62); all three contain the F and G proteins. They differ in the incorporation of the DNA pilot and internal scaffolding proteins. The presence or absence of the B protein, a substrate for the ompT protease, is an artifact of purification procedures (105). However, ompT cleavage is not required for morphogenesis (32). Although the incorporation of the H protein remains obscure, its absence does not affect the formation of capsid-like structures in vivo (123).

The 12S particle exhibits the biochemical properties traditionally associated with morphogenetic intermediates: the ability to be chased into large particles in temperature-shift experiments with *tsD* mutants. However, it is not clear whether this particle represents a true morphogenetic intermediate or the product of an off-pathway but reversible reaction (62). The procapsid atomic structure supports the latter possibility. However, the particle matured during crystallization and most likely does not represent the biologically significant intermediate. There are apparently few contacts made between the F and G proteins within the structure. The spike pentamers are tethered to the underlying capsid proteins via the external scaffolding protein, which is not a component of the 12S particle. Yet the 12S particle formed in *tsD* infections is most likely not the degradation product of a fully formed procapsid. Such degradation products lose the ability to re-enter the morphogenetic pathway. For example, in cells infected with *csD* alleles at restrictive tempecratures, fragile procapsids are produced. During DNA packaging, these procapsids dissociate into 12S-like particles that cannot be chased into larger structures (42). Then what is the true assembly intermediate? The atomic structure suggests that it should contain not only the internal scaffolding, major spike, and capsid proteins but also 20 copies of the external scaffolding protein. This may be the fleeting 18S particle (28, 45).

The External Scaffolding Protein

The ϕX174 external scaffolding protein (protein D) performs many of the functions typically associated with internal species in systems with one scaffolding protein: the

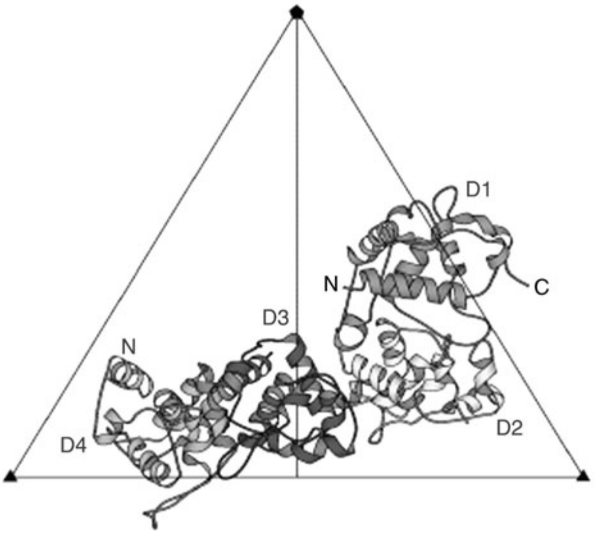


Figure 11-3 The structures of the external scaffolding protein.

organization of assembly precursors into a procapsid and the stabilization of that structure. However, its function is physically and temporally dependent on the internal scaffolding protein, which induces the conformational changes in capsid pentamers. In the procapsid crystal structure, 20 D proteins are associated with each pentameric capsomer (35, 36). Remarkably, there is little or no contact between capsid pentamers. The structure is held together primarily by 2-fold related contacts between D proteins. The four D subunits (D1, D2, D3, and D4) per asymmetric unit (figure 11-3) are arranged as two similar, but not identical, asymmetric dimers (D1D2 and D3D4). These dimeric subassemblies may be a component of the tetrameric structures seen in solution (R. McKenna, personal communication). However, these particles have closed point group symmetry, a significant departure from the appearance of the D subunits in the asymmetric unit. It has not been determined whether these tetrads consist of four or eight proteins. These particles sediment at 4S and behave like assembly intermediates in pulse-chase experiments (132).

The "canonical monomer" in the crystal structure is composed of seven α-helices separated by loop regions. However, there is considerable structural variation between the subunits, which bear no resemblance, not even an inkling, to quasi-equivalence. For example, α-helix #5′ forms a β structure in subunit D2, where it participates in interdimer contacts with D3, and a helical structure in D3, where it participates in D4 intradimer contacts. In subunit D4, it mediates scaffolding contacts across the 2-fold axis of symmetry and contains no secondary structure. α-helix #7 only forms in the D4 subunit, where it mediates the most extensive coat protein interactions found in the entire lattice. Structural data also suggest that this unique arrangement is mediated, in part, by glycine residue 61 in α-helix #3. One monomer in each dimer is bent 30° at this site. The kink may be needed to switch the second monomer into a non-sticky conformation. The importance of this residue is even emphasized in the structure of the genome. The glycine 61 and gene E start codons overlap in all the ϕX174-like phages (52, 83, 107). Hence there is strong selective pressure to maintain the codon. A limited number of site-directed mutants can be and have been generated at this site (27). The mutants can only be propagated in cells overexpressing the wild-type protein. In coinfections with wild-type ϕX174, the lethal phenotype is dominant, suggesting morphogenetic defects occurring after monomers interact to form dimers.

While the atomic structure of the ϕX174 procapsid has yielded a wealth of information, it does not reveal all the contacts in which the external scaffolding protein participates. During crystallization, the procapsid matured: the coat protein assumed a conformation similar to its structure in the mature virion. There are two dramatic differences between the X-ray and the cryo-electron microscopy models. First, in the X-ray model, the coat protein has moved inward radially, away from the external scaffolding protein lattice. However, the coat protein has not fully dissociated from it. Second, a large coat protein α-helix occupies the 3-fold axes of symmetry, which in the cryo-electron microscopic image is free of density and thus contains a 3 nm pore. In addition, the interactions between the procapsid and the packaging machinery would not be visualized. The results of genetic experiments and experiments conducted with inhibitory cross-species and chimeric scaffolding proteins have elucidated some of these unseen contacts. The external scaffolding protein primary sequences of the ϕX174 and the related bacteriophage α3 are 70% conserved (83, 107). Divergent sequences are localized to the N- and C-termini of the proteins. These sequences constitute α-helices #1, #7, and loop #6 in the atomic structure. The ϕX174 and α3 external scaffolding genes have been cloned and expressed in vivo and do not cross-complement. However, expression of a *foreign* scaffolding protein blocks wild-type morphogenesis (26). The ability of foreign scaffolding proteins to inhibit morphogenesis is most likely due to the formation of cross-species dimers. The regions of the proteins responsible for dimerization (α-helices 2–6) are strongly conserved.

To determine whether one or both termini confer inhibitory effects, chimeric genes, in which the first α-helices were interchanged, have been constructed and expressed in vivo. Expression inhibits morphogenesis in a somewhat species-specific manner. Efficient inhibition is governed by the identity of the first α-helix. The chimera that contains α-helix #1 from ϕX174, for example, strongly inhibits ϕX174 morphogenesis; α3 morphogenesis is only modestly affected. The relative levels and species specificity of inhibition merit further explanation. While the phenomenon was symmetrical—depending on which phage, ϕX174 or α3, was used in the assay—for clarity the discussion will focus exclusively on ϕX174. The weak inhibition conferred by the α3/ϕX chimera is only achieved when the chimeric gene is maximally induced. Under these conditions, inhibition in plating assays ranged from 10^{-1} to 10^{-3}. The strong inhibitory phenomenon is observed when the cloned ϕX/α3 D gene is barely induced. Even under those conditions, plating efficiencies drop below 10^{-6}. Although plating efficiencies below 10^{-6} are also achieved when expressing the wild-type foreign α3 protein, the cloned gene must be maximally induced. These data suggest a temporal mechanism in which the initial recognition of the coat protein is mediated by α-helix #1 of the external scaffolding protein. The presence of the proper first α-helix, of the same origin as the viral coat protein, facilitates the incorporation of the chimeric protein into external lattice, acting as a vehicle for the incorporation of inhibitory foreign loop #6 and α-helix #7 sequences.

ϕX174 intermediates synthesized in cells expressing foreign or the chimeric ϕX/α3 D proteins were analyzed by sucrose gradient sedimentation. In extracts generated from cells expressing the ϕX/α3 D chimera, which assays for defects conferred by foreign loop #6 and α-helix #7 structures, procapsids and empty capsids were present; however, virions were not, indicating a block in DNA packaging. The docking of the replication/packaging machinery is most likely prevented by sequences found in the C-terminus of the α3 protein. In addition, ϕX174 mutants resistant to the expression of ϕX/α3 D chimera have been isolated (*chiD*^{R} mutants). These mutations alter viral protein A, a component of the genome biosynthesis/packaging machinery, which binds the procapsid during DNA packaging, presumably at the 2-fold axis of symmetry (43). In extracts prepared from cells expressing the foreign protein, procapsids (108S) and empty capsids (70S) were not detected, suggesting either a block before procapsid formation or the production of unstable particles. This suggests that all the external scaffolding dimers were heterogeneous and some of them, keeping in mind that D proteins form asymmetric dimers, may not have been able to recognize the coat and/or spike proteins due to the presence of the foreign α-helix sequence (26). Interactions between α-helix #1 of the D1 subunit and the spike protein are visible in the X-ray model. The helix's proximity to the 3-fold axis of symmetry in the D4 subunit also suggests an interaction with the coat protein. However, this interaction cannot be observed in the atomic structure due to the above-mentioned maturation events that occurred during crystallization.

In complementation experiments, neither the foreign loop #6 nor helix #7 chimeras, built directly into the viral genome, complemented the α-helix #1 chimera. This may indicate that α-helix #1 also plays a critical role in the D4 subunit. However, an element of uncertainty exists in intragenic complementation experiments, due to the lack of a positive control. Genetic data (42, 46) also suggest an unseen interaction occurs between D4 α-helix #1 and α-helix #4 of the viral coat protein. Both helices are found at the 3-fold axis of symmetry in the closed structure, which has matured during crystallization. Two point mutations in the first α-helix of protein D have been extensively characterized (42, 46). These mutations confer a fragile procapsid phenotype. While procapsid morphogenesis is not inhibited, the particles disassociate during DNA packaging. In an open structure, the coat protein helix could be shifted upwards and may contact α-helix #1 of the D4 subunit of an adjacent asymmetric unit, which is the most closely associated subunit with the underlying coat protein. Both helices are amphipathic and could interact via hydrophobic interfaces. If this interaction is indeed present in the native structure, this may exclude dimers with a foreign α-helix #1 from the D3D4 position. In addition, results of experiments conducted with chimeric G4/ϕX174 external scaffolding proteins and mutant ϕX174 coat proteins capable of productively utilizing the chimeric scaffolding also suggest interactions between the D protein's first α-helix and 3-fold related coat protein residues (B. A. Fane, unpublished data).

To further dissect structure–function relationships in the ϕX174 external scaffolding protein, additional chimeras have been generated in the C-terminus of the protein. The chimeric genes were built directly into the phage genome and substitute either α3 loop #6 or helix #7 sequences into the ϕX174 gene. The helix #7 chimera can be propagated in cells overexpressing the wild-type protein. However, in coinfections with wild-type ϕX174, the chimera confers a dominant lethal phenotype. The severity of dominant phenotype is greatly decreased in coinfections with the *chiD*^{R} strain, indicating that the *chiD*^{R} mutation confers resistance to the defects caused by the incorporation of this inhibitory region. The loop #6 chimera, on the other hand, is viable, but has a cold-sensitive (*cs*) phenotype. Both extragenic and intragenic revertants of the *cs* phenotype have been isolated (27). The intragenic mutation is in loop #6, conferring an E → D substitution in the central residue of the loop, which is the amino acid found in the ϕX174 sequence. From the atomic structure of the procapsid, it is known that this residue, but only in the D4 subunit, makes extensive contacts with a lysine 118 of the coat protein. However, ϕX174 morphogenesis is a dynamic system, and it is difficult to conceive successful evolution if morphogenesis is contingent on the identity of one amino acid. Hence the extragenic suppressors are likely of equal import. These mutations confer amino acid substitutions on the outer surface of the capsid protein, but they are not limited to the contacts made by the D4 subunit. This suggests that the D4 position can be adjusted by neighboring D proteins within the lattice. The identification of these suppressors serves another important role: elucidating interactions lost in the X-ray model due to particle maturation within the crystal. This maturation includes the radial collapse of capsid proteins away from the external scaffolding lattice (14, 36). While the X-ray structure reveals coat protein contacts for each D subunit, they are not extensive. The extragenic suppressors map to residues adjacent to the known contacts, indicating that the coat–external scaffolding interface in the native structure is more extensive than the X-ray model reveals.

The relative ability of the chimeric proteins to inhibit morphogenesis suggests a temporal model for coat–external scaffolding protein recognition. The cloned α3/ϕX chimeric protein only weakly inhibits ϕX174 morphogenesis. In contrast, the ϕX/α3 chimera is a potent inhibitor. These data suggest that chimeric protein incorporation into the lattice is a function of the first α-helix. Its proximity to the 3-fold axis of symmetry and genetically defined interactions with the viral coat protein (42, 46) suggest that this first substrate-specific interaction occurs between a dimer of scaffolding protein and the adjacent 5-fold related coat protein. Figure 11-4 illustrates the atomic structure of the 3-fold axis of symmetry in the ϕX174 procapsid. In the native structure, this axis would not be occupied by the large coat protein helix. This helix may be restrained by the D4 subunit of the adjacent asymmetric unit. After coat–scaffolding recognition, loop #6 and α-helix #7 place the D4 subunit dimer atop the capsid. D1D2 dimers would then be added to the same asymmetric unit and the adjacent unit, mediated by 5-fold D4–D2 and D4–D1 interactions, respectively. In a chain reaction, dimers would add around the pentamer. The resulting intermediate would contain five copies of the spike, coat, and internal scaffolding proteins, one copy of the minor spike protein H, and 20 copies of protein D. A particle of this composition has been detected in vivo and it sediments at 18S (28, 45). However, a means to genetically trap this intermediate in large quantities has not been established.

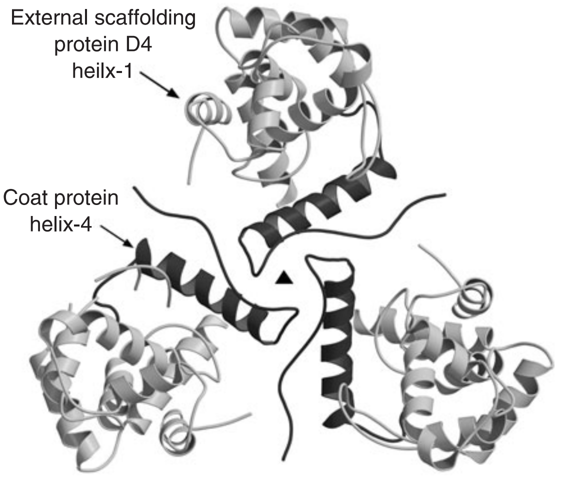


Figure 11-4 The 3-fold axes of symmetry of ϕX174 "closed" procapsid. In the native structure, the large coat helix does not sit occupy the 3-fold axes of symmetry, which contain pores. The results of genetic analyses suggest interactions between the first α-helix of the D4 subunit and the large coat protein helix.

In a kinetic model of capsid assembly, as described for P22 (103), procapsid formation would be nucleated by a rate-limiting step, and hence a higher order reaction than those which follow. If post-nucleation morphogenesis involves the addition of one pentameric intermediate to a growing shell, the nucleation complex formation would require at least three pentameric intermediates. Since there are no coat–coat or 3-fold related scaffolding contacts in the procapsid atomic structure, the reaction is expected to be catalyzed by three sets of 2-fold related interactions. The involvement of the external scaffolding protein is easily visualized, due to the specific and ordered 2-fold related contacts in the crystal structure. The role, if any, of the internal scaffolding protein in procapsid nucleation remains more obscure.

DNA Packaging and the DNA Binding Protein

Genome biosynthesis and packaging are concurrent processes. The pre-initiation complex, consisting of the host cell *rep*, viral A and C proteins, associates with the procapsid forming the 50S complex (95). The viral A protein binds the origin of replication in replicative-form DNA, as described above. This is both necessary and sufficient for packaging specificity (3, 49). The location of the pre-initiation docking site has been elucidated in a series of genetic experiments (43). ϕX174 morphogenesis was examined in two hosts with mutations in the bacterial *rep* gene (128). Morphogenesis was blocked at the formation of the 50S complex. Stage II DNA synthesis, which also requires the bacterial *rep* protein, was not affected. Procapsids accumulate in these infected cells. A second-site genetic analysis was conducted. Second-site phage-encoded mutations were isolated in the genes encoding the viral coat and A proteins. The mutations within the coat protein cluster in the atomic structure of the virion, tracing a pronounced depression that skirts the 2-fold axis of symmetry. The location of these substitutions suggests that the depression may serve as the pre-initiation complex binding site during 50S particle formation. Both genetic and structural data support this hypothesis. The second site suppressors are active in *trans* and confer dominant phenotypes.

The DNA binding protein, protein J, enters the procapsid during packaging (50) and is absolutely required for genome encapsidation (57, 58). Furthermore it may dislodge the internal scaffolding protein from a common binding cleft in the coat protein (35, 36, 92–94). Protein J is a small, extremely basic protein, 37 amino acids in length and separated into four functional domains. It binds the genome via simple charge–charge interactions (81) mediated by two DNA binding domains. These domains contain the 12 basic residues of the protein and are separated by a short sequence rich in proline residues. Once in the procapsid, the C-terminus of the protein, which is very hydrophobic and aromatic, binds to a coat protein cleft (93, 94). This may facilitate further interactions between the genome and capsid, specifically with a small cluster of adjacent basic capsid amino acids (93, 94).

Genetic studies with the ϕX174 J protein have demonstrated that the ability to bind and package DNA is directly related to the number of basic amino acid residues in the protein (56, 81). Step-wise substitutions for the basic residues produce less functional proteins. The removal of fewer than three basic residues produces viable particles. However, the particles often have *cs* and *ts* phenotypes and very small plaque morphologies. Substitutions of three or four basic residues produce fully packaged particles that have lost infectivity. These mutants have dominant lethal phenotypes in coinfections with wild-type ϕX174. The dominant phenotype and the production of packaged particles indicate that the mutant proteins retain enough function to enter into the morphogenetic pathway. Substitutions of six basic amino acids confer recessive lethal phenotypes. However, the proteins retain a low level of function. Stage III DNA synthesis is completed and unit-length genomes are associated with capsids, but are not fully encapsidated, remaining sensitive to DNase (56).

Genome–Capsid Interactions and the Final Collapse

Unlike large dsDNA bacteriophages (38), the ϕX174 genome does not exist as a dense core in the capsid. Instead, the DNA binding protein and basic capsid amino acid residues tether it to the capsid's inner surface. There are 60 copies of protein J per virion, one associated with each coat protein. In the atomic model, the protein forms an S-shaped polypeptide chain devoid of secondary structure. The C-terminus of the protein is tightly associated with a cleft, located near the center of the coat protein. Moving toward the N-terminus, the protein traces a path toward the 5-fold axis of symmetry, crosses over to the adjacent capsid protein, and veers toward the C-terminus of the adjacent J protein. This motif suggests that the DNA binding protein guides the incoming genome into a somewhat ordered conformation. Accordingly, a portion of the genome is ordered in the X-ray structure (93, 94).

The biophysical characterization of fully packaged particles with mutant DNA binding proteins suggest that the ϕX174 genome, more specifically its interactions with the DNA binding and capsid proteins, may also perform a morphogenetic function, mediating the final stages of morphogenesis (56). These final stages involve the dissociation of the external scaffolding protein and a 0.85 nm radial collapse of capsid pentamers (14, 35, 36, 72). The mutant particles were significantly denser than wild-type, but the protein composition of the two particles appears to be identical. Therefore, gross excesses of protein J within the volumetrically fixed capsid do not cause the altered density. The effects of possible Cs^+ permeability were more difficult to discern. However, extragenic second-site suppressor procapsids packaged with the mutant DNA binding protein restored particle densities to near wild-type values. Therefore, a model in which counter-ions compensate for the loss of basic amino acids is not the basis of the density differences. Differences are also expressed on the capsid exterior. In host cell attachment assays, mutant particles exhibited substantially lower attachment efficiencies. Dramatic differences in native gel migration, which is a function of size and net surface charge (112), were also observed. In all assays, the extragenic suppressor restored the properties of the mutant particles to nearly those of wild-type virions. Naked ϕX174 DNA is substantially richer in secondary structure than packaged DNA (13). Therefore, if an interplay between base-pairing and DNA–capsid association occurs, altering the base composition of the genome should also produce altered particles. ϕX174 ampicillin transducing particles were generated by packaging single-stranded versions of unit-length plasmids. As with the particles packaged with the mutant DNA binding proteins, the transducing particles exhibited different biophysical characteristics from wild-type ϕX174.

The role of DNA–capsid interactions in ϕX174 is obviously not as dramatic as those seen in other viral systems such as Flock House, Southern cowpea mosaic virus or Brome mosaic virus. In those systems abrogating genome–capsid interactions leads to either polymorphic particles or capsids with altered T values (37, 86, 108). In the procapsid there are no discernible pentamer–pentamer interactions. The integrity of the capsid appears to be maintained by the scaffolding proteins. After packaging, the internal scaffolding protein is extruded from the structure and replaced by the DNA binding protein and the tethered genome. This may supplant scaffolding function in the provirion. The provirion-to-virion transition is marked by the release of the external scaffolding protein and the completion of the 0.85 nm radial collapse of coat protein pentamers. This tether constrains the spatial orientation and secondary structure of the remaining nucleotides (13). Therefore altering the tether or the base composition of the packaged nucleic acid may affect the magnitude or the integrity of the collapse.

Lysis

A more detailed discussion of bacteriophage lysis strategies, including the *Microviridae*, can be found in chapter 10. Therefore, it will only be summarized here. Unlike large double-stranded bacteriophages that encode a two-component lysis system, comprising an endolysin and a holin, small bacteriophages, such as the *Microviridae*, do not have the genetic capacity to encode a two-component system (140). To circumvent this constraint, the *Microviridae*, as well as other small bacteriophages, have evolved a small protein that inhibits a host cell enzyme involved in peptidoglycan biosynthesis (15). In the *Microviridae*, the E protein performs this function (70). Once translated, the protein is found associated with the cell membrane (2). The mechanism for E protein action has been controversial for many years. However, the results of recent genetic experiments have elucidated the mechanism of this "antibiotic-like" protein.

The expression of an inducible cloned E gene leads to cell lysis (139). Therefore E protein is both necessary and sufficient for this function. However, mutant cells resistant to gene E expression, *slyD* (*s*ensitivity to *ly*sis), can be easily isolated (90). The phenotype is conferred by mutations in a peptidly-prolyl *cis*-transferase-isomerase, or PPIase (106). However, it is unlikely that the *slyD* gene product is the E protein target. Considering the function of PPIases, protein folding (33), and the observation that protein E does not associate with the membranes in *slyD* cells, it is more likely that E protein, with five prolyl bonds, is a PPIase substrate. ϕX174 gene E mutants, *Epos* (*p*lates *o*n *slyD*), were also isolated. *Epos* proteins were expressed in an *E. coli slyD* host to identify other host cell genes that confer a lysis-resistant phenotype (15). The surviving colonies contained mutations in the *mraY* gene, which encodes translocase I. This enzyme catalyzes the formation of the first lipid-linked intermediate in cell wall biosynthesis (71). The results of both in vitro and in vivo studies strongly suggest that E protein-specifically inhibits translocase I catalyzed reaction (16). The lysis of ϕX174-infected cells requires cell growth (19). The similarities between E-protein-mediated lysis and penicillin-mediated lysis are apparent.

Evolution and Evolutionary Studies

In the past decade, two approaches have been taken to investigate *Microviridae* evolution. In one approach, pioneered by the Drs. J. J. Bull and H. A. Wichman and colleagues, viruses are placed under selective conditions, either high temperature or host variations, and grown for numerous generations in a chemostat (23, 24, 31, 69, 137). At various time intervals, individual genomes are sequenced. Therefore the appearance and disappearance of beneficial mutations can be monitored. Of course, host adaptation experiments have identified mutations that increase attachment and penetration efficiency. But mutations that affect intracellular tropism, more specifically mutations in gene A, were also uncovered. Protein A adaptation would be needed to optimize its interaction with the host cell *rep* protein, which functions as a host cell helicase during stage II and stage III replication (41). The high temperature selection yielded several types of mutations. In addition to mutations that appear to affect both intracellular and extracellular interactions, which may be needed to stabilize macromolecular interactions at higher temperatures, many mutations affecting morphogenesis and/or the stability of the procapsid were isolated. Considering the metastable properties of assembly intermediates, these results are not surprising.

In both selections, mutations in genetic regulatory sequences were also recovered. These mutations most likely optimize the relative level of viral proteins synthesized under the experimental conditions. However, it should not be assumed that these mutations lead to elevations in transcription, transcript stability, or translation. Adaptation may involve maintaining an optimal balance of viral components (48, 125). For example, the expression of a relatively stable protein under selection conditions may be downregulated while the expression of less stable proteins may be upregulated. Some recurrently recovered neutral mutations may also be acting on this level, changing the intracellular level of the encoded protein by codon usage. However, neutral mutations may also be acting on a structural level by altering a genome secondary structure. As discussed above, the interplay between a genome's secondary structure and its interactions with the capsid's inner surface may affect the final stages of virion morphogenesis and can create capsid surface distortions that reduce host cell attachment. In several the selections, a deletion in the F–J intercistronic group was recovered. This deletion was also isolated as a mutation that elevates the rate of host cell attachment and penetration (74). In addition, it acts as a second-site suppressor of mutations that affect the interface between the genome and the inner surface of the viral capsid (S. Hafenstein and B. A. Fane, unpublished data).

The second area of evolutionary research has focused on the isolation of novel members of the family. As stated by Hendrix et al. (66, 67), the prevalence of double-stranded DNA phages and prophages—cryptic, defective, and replication competent—creates an enormous pool of evolutionary material which can be horizontally exchanged, otherwise known as the moron accretion hypothesis. Consequently, a mosaic spectrum of related phage species has arisen. In contrast, the members of the *Microviridae* appear to fall into two distinct and rather distantly related subfamilies. Protein homologies between the two subfamilies are approximately 20% or less (table 11-2), a typical value when comparing the most distantly related members of either the lambda or T4-like groups (66, 67, 130). However unlike tailed dsDNA families, no mosaic species that bridge the evolutionary chasms have been isolated. The members of the ϕX174 subfamily were isolated from γ-proteobacteria,

Table 11-2 Amino Acid Identities of ϕMH2K Gene Products with Chp1, Chp2, SpV4, and ϕX174, and Amino Acid Homologies

Gene[a] product	Percent amino acid identity				
	Chp2	Chp1	SpV4	Chp2/Chp1[b]	ϕX174-like[c]
VP1	46.9	40.4	38	49.6	19 (α3 F)
VP2	26.5	21.3	25	29.9	20 (α3 H)
VP3	32	27.6	18.4	27.3	18 (α3 B)
VP4	27.9	22.5	27	22.2	18 (G4 A)
VP5	39.5	26.8	18.4	30.2	20 (α3 C)
VP8	32.6	33.3	31	54.5	21 (G4 J)

[a] Genes (and gene products) which are conserved between ϕMH2K and the chlamydiaphages.
[b] Comparison between Chp1 and Chp2 proteins.
[c] Comparisons between ϕX174, G4 or α3.

and are fairly closely related. In contrast, the members of the second subfamily were isolated from a very diverse group of hosts: *Chlamydia*, *Bdellovibrio* (δ-proteobacteria), and *Spiroplasma* (20, 88, 104, 126). Despite the extreme diversity of these hosts, the viruses within the subfamily are remarkably similar. In fact, the *Bdellovibrio* virus ϕMH2K is more closely related to some of the *Chlamydia* viruses, than the *Chlamydia* viruses are related to each other (table 11-2), suggesting that species jumping may play a major role in the evolution of this family.

There are several factors that may limit horizontal transfer in ssDNA viruses, such as small circular genomes and lytic life cycles that do not require recombination. In addition, the small T = 1 capsid may restrict the incorporation of exogenous DNA sequences, or morons (66, 67). Since it is unlikely that these small viruses can acquire morons, all members of the *Microviridae* appear to possess preserved open reading frames, found mostly in overlapping genes. Mutations could accrete in these reading frames, or cretins, until a gene encoding a beneficial function is produced. Examples of ϕX174 cretins may include lysis gene E, and genes A* and K, both unessential and of unclear function (30, 129).

Genetic maps of ϕMH2K, ϕX174, and Chp2 are given in figure 11-5. Neither Chp2 nor ϕMH2K encodes an external scaffolding or major spike protein, ϕX174 D and G proteins, respectively. The loss of these genes accounts for the smaller ϕMH2K and Chp2 genomes. The external scaffolding protein has at least two known functions in ϕX174 morphogenesis. It stabilizes the procapsids at the 2- and 3-fold axes of symmetry and directs the placement of the major spike protein (35, 36). These functions are either not required or performed by different proteins in the ϕMH2K-like phages. First, there is no major spike protein. The 2-fold stabilization function may be performed by Vp3, the internal scaffolding protein equivalent, which in ϕX174 self-associates across 2-fold axes of symmetry. Finally, as seen in the cryo-image reconstruction of SpV4, a large coat protein insertion loop forms spikes at the 3-fold axis of symmetry (29). This large insertion loop may be the relic of the ancestral external scaffolding or the major spike protein. Coding Vp3 in a normal reading frame, as opposed

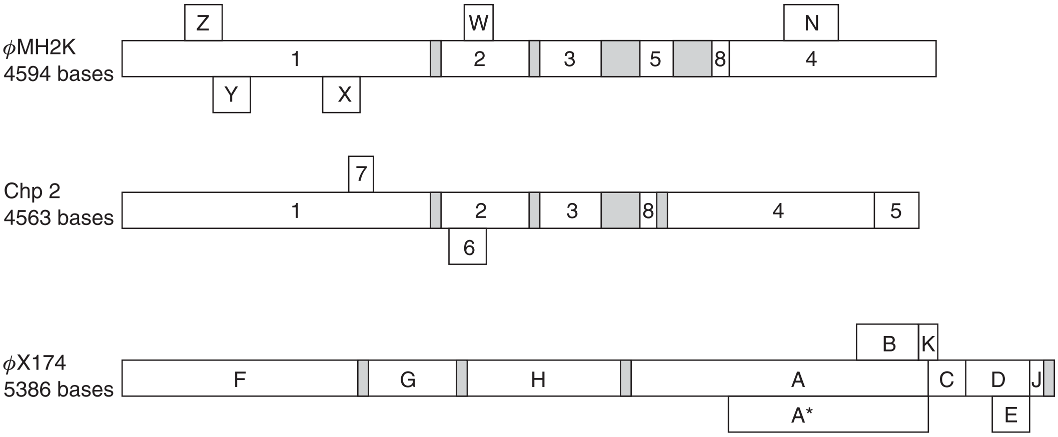


Figure 11-5 The genetic maps ϕMH2K, Chp2, and ϕX174 phages. Reading frames in Chp2 and ϕMH2K that encode homologous proteins have the same gene numbers. Chp2 open reading frames (ORFs) 6 and 7, and ϕMH2K ORFs W, X, Y, Z, and N are cretins.

to a cretin, may be a related phenomenon. The B proteins of the ϕX174-like phages are highly diverged, yet they cross-function, suggesting that interactions are primarily nonspecific and flexible (25, 28). With the loss of the external scaffolding protein, internal scaffolding protein interactions may need to be more specific, requiring a reading frame unconstrained by other genes.

Future Prospects

Considering the history of *Microviridae* research, it is difficult to predict its future. For example, by the late 1980s, *Microviridae* research, so popular in the 1960s and 1970s, was being studied by only a few research groups. Then came the atomic structure of the virion and procapsid, the discovery of the ϕMH2K subfamily, nanotechnology and the ability to make transgenic animals. Animals transgenic for the ϕX174 genome are being developed for eukaryotic mutagenesis experiments (133). With the elucidation of X-ray models, genetic maps of biochemical phenotypes transubstantiate from straight lines to functional domains, and the domains have structures. The consequences for morphogenetic and evolutionary research are apparent. The ϕMH2K-Chp2-like viruses could lead to the first DNA transfer system for *Chlamydia*, opening areas of molecular research which have been hindered by the lack of a laboratory-based genetic system. The atomic structure of the ϕ29 connector represents a powerful nano-motor that must drive the packaging of DNA to the concentration of liquid crystal (116). Considering protein–DNA interactions found in the *Microviridae*, is the pentameric spike protein a component of the motor that is required to force the DNA out of the capsid? O brave new world that hath such particles in't!

Acknowledgments

The authors would like to acknowledge the support of the National Science Foundation and of their colleagues and collaborators who freely exchange ideas, host cells, and viral mutants. The open and helpful interactions between these various groups make *Microviridae* research more productive, stimulating, and enjoyable.

References

1. Air, G. M., A. R. Coulson, J. C. Fiddes, T. Friedmann, C. A. Hutchison III, F. Sanger, P. M. Slocombe, and A. J. Smith. 1978. Nucleotide sequence of the F protein coding region of bacteriophage phiX174 and the amino acid sequence of its product. J. Mol. Biol. 125:247–254.
2. Altman, E., K. Young, J. Garrett, R. Altman, and R. Young. 1985. Subcellular localization of lethal lysis proteins of bacteriophages lambda and phiX174. J. Virol. 53:1008–1111.
3. Aoyama, A., and M. Hayashi. 1985. In vitro packaging of plasmid DNAs into ϕX174 bacteriophage capsid. Nature 297:704–707.
4. Aoyama, A., and M. Hayashi. 1986. Synthesis of bacteriophage ϕX174 in vitro: mechanism of switch from DNA replication to DNA packaging. Cell 47:99–106.
5. Arai, K., and A. Kornberg. 1981. Unique primed start of phage ϕX174 DNA replication and mobility of the primosome in a direction opposite chain synthesis. Proc. Natl. Acad. Sci. USA 78:69–73.
6. Arai, K., R. Low, J. Kobori, J. Shlomai, and A. Kornberg. 1981. Mechanism of dnaB protein action. V. Association of dnaB protein, protein n′, and other prepriming proteins in the primosome of DNA replication. J. Biol. Chem. 256:5273–5281.
7. Arai, K., R. McMacken, S. Yasuda, and A. Kornberg. 1981. Purification and properties of *Escherichia coli* protein i, a prepriming protein in ϕX174 DNA replication. J. Biol. Chem. 256:5281–5287.
8. Axelrod, N. 1976. Transcription of bacteriophage ϕX174 in vitro: analysis with restriction enzymes. J. Mol. Biol. 108:771–779.
9. Axelrod, N. 1976. Transcription of bacteriophage ϕX174 in vitro: selective initiation with oligonucleotides. J. Mol. Biol. 108:753–770.
10. Baas, P. D., W. R. Teertstra, A. D. van Mansfeld, H. S. Jansz, G. A. van der Marel, G. H. Veeneman, and J. H. van Boom. 1981. Construction of viable and lethal mutations in the origin of bacteriophage ϕX174 using synthetic oligodeoxyribonucleotides. J. Mol. Biol. 152:615–639.
11. Bayer, M. E., and T. W. Starkey. 1972. The adsorption of bacteriophage ϕX174 and its interaction with *Escherichia coli*: a kinetic and morphological study. Virology 49:236–256.
12. Benbow, R. M., C. A. Hutchison III, J. D. Fabricant, and R. L. Sinsheimer. 1971. Genetic map of bacteriophage ϕX174. J. Virol. 7:549–558.
13. Benevides, J. M., P. L. Stow, L. L. Ilag, N. L. Incardona, and G. J. Thomas, Jr. 1991. Differences in secondary structure between packaged and unpackaged single-stranded DNA of bacteriophage ϕX174 determined by Raman spectroscopy: a model for ϕX174 DNA packaging. Biochemistry 30:4855–4862.
14. Bernal, R. A., S. L. Hafenstein, N. H. Olson, V. D. Bowman, P. R. Chipman, T. S. Baker, B. A. Fane and M. G. Rossmann. 2003. Structural studies of bacteriophage alpha3 assembly. J. Mol. Biol. 325:11–24.
15. Bernhardt, T. G, W. D. Roof, and R. Young. 2000. Genetic evidence that the bacteriophage phi X174 lysis protein inhibits cell wall synthesis. Proc. Natl. Acad. Sci. USA 97:4297–4302.
16. Bernhardt, T. G., D. K. Struck, and R. Young. 2001. The lysis protein E of ϕX174 is a specific inhibitor of the mraY-catalyzed step in peptidoglycan biosynthesis. J. Biol. Chem. 276:6093–6097.
17. Blasi, U., K. Nam, W. Lubitz, and R. Young. 1990. Translational efficiency of ϕX174 lysis gene E is unaffected by upstream translation of the overlapping gene D reading frame. J. Bacteriol. 172:5617–5623.

18. Blattner, F. R., G. Plunkett III, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau and Y. Shao. 1997. The complete genome sequence of *Escherichia coli* K-12. Science 277:1453–1474.
19. Bradley, D. E., C. A. Dewar, and D. Robertson. 1969. Structural changes in *Escherichia coli* infected with a ϕX174 type bacteriophage. J. Gen. Virol. 5:113–121.
20. Brentlinger, K., S. Hafenstein, C. R. Novak, B. A. Fane, R. Birgon, R. McKenna, and M. Agbandje-McKenna. 2002. *Microviridae*, a family divided. Isolation, characterization and genome sequence of a ϕMH2K, a bacteriophage of the obligate intracellular parasitic bacterium *Bdellovibrio bacteriovorus*. J. Bacteriol. 184:1089–1094.
21. Brown, D. T., J. M. MacKenzie, and M. E. Bayer. 1971. Mode of host cell penetration by bacteriophage ϕX174. J. Virol. 7:836–846.
22. Brown, D. R., M. J. Roth, D. Reinberg, and J. Hurwitz. 1984. Analysis of bacteriophage ϕX174 gene A protein mediated termination and reinitiation of ϕX174 DNA synthesis. I. Characterization of the termination and reinitiation reactions. J. Biol. Chem. 259:10545–10555.
23. Bull, J. J., Badgett, M. R., and H. A. Wichman. 2000. Big-benefit mutations in a bacteriophage inhibited with heat. Mol Biol. Evol. 17:942–950.
24. Bull, J. J., M. R. Badgett, H. A. Wichman, J. P. Huelsenbeck, D. M. Hillis, A. Gulati, C. Ho, and I. J. Molineux. 1997. Exceptional convergent evolution in a virus. Genetics 147:1497–1507.
25. Burch, A. D., and B. A. Fane. 2000. Efficient complementation by chimeric *Microviridae* internal scaffolding protein is a function of the COOH-terminus of the encoded protein. Virology 270:286–290.
26. Burch, A. D., and B. A. Fane. 2000. Foreign and chimeric external scaffolding proteins as inhibitors of *Microviridae* morphogenesis. J. Virol. 74:9347–9352.
27. Burch, A. D., and B. A. Fane. 2003. Genetic analyses of putative conformation switching and cross-species inhibitory domains in *Microviridae* external scaffolding proteins. Virology 310:64–71.
28. Burch, A. D., J. Ta, and B. A. Fane. 1999. Cross-functional analysis of the *Microviridae* internal scaffolding protein. J. Mol. Biol. 286:95–104.
29. Chipman, P. R., M. Agbandje-McKenna, J. Renaudin, T. S. Baker, and R. McKenna. 1998. Structural analysis of the Spiroplasma virus, SpV4: implications for evolutionary variation to obtain host diversity among the *Microviridae*. Structure 6:135–145.
30. Colasanti, J., and D. T. Denhardt. 1987. Mechanism of replication of bacteriophage ϕX174. XXII. Site-specific mutagenesis of the A* gene reveals that A* protein is not essential for ϕX174 DNA replication. J. Mol. Biol. 197:47–54.
31. Crill, W. D., H. A. Wichman, and J. J. Bull. 2000. Evolutionary reversals during viral adaptation to alternating hosts. Genetics 154:27–37.
32. Dalphin, M. E., B. A. Fane, M. O. Skidmore, and M. Hayashi. 1992. Proteolysis of bacteriophage phi X174 prohead accessory protein gpB by *Escherichia coli* OmpT protease is not essential for phage maturation in vivo. J. Bacteriol. 174:2404–2406.
33. Deuerling, E., A. Schulze-Specking, T. Tomoyasu, A. Mogk, and B. Bukau. 1999. Trigger factor and DnaK cooperate in folding of newly synthesized proteins. Nature 400:693–696.
34. Ding, Y., R. L. Duda, R. W. Hendrix, and J. M. Rosenberg. 1995. Complexes between chaperonin GroEL and the capsid protein of bacteriophage HK97. Biochemistry 34:14918–14931.
35. Dokland, T., R. A. Bernal, A. Burch, S. Pletnev, B. A. Fane, and M. G. Rossmann. 1999. The role of scaffolding proteins in the assembly of the small, single-stranded DNA virus ϕX174. J. Mol. Biol. 288:595–608.
36. Dokland, T., R. McKenna, L. L. Ilag, B. R. Bowman, N. L. Incardona, B. A. Fane, and M. G. Rossmann. 1997. Structure of a viral procapsid with molecular scaffolding. Nature 389:308–313.
37. Dong, X. F., P. Natarajan, M. Tihova, J. E. Johnson, and A. Schneemann. 1998. Particle polymorphism caused by deletion of a peptide molecular switch in a quasiequivalent icosahedral virus. J. Virol. 72:6024–6033.
38. Earnshaw, W. C., and S. R. Casjens. 1980. DNA packaging by the double-stranded DNA bacteriophages. Cell 21:319–331.
39. Eisenberg, S., J. Griffith, and A. Kornberg. 1977. ϕX174 cistron A protein is a multifunctional enzyme in DNA replication. Proc. Natl. Acad. Sci. USA 74:3198–3202.
40. Eisenberg, S., and A. Kornberg. 1979. Purification and characterization of ϕX174 gene A protein. A multifunctional enzyme of duplex DNA replication. J. Biol. Chem. 254:5328–5332.
41. Eisenberg, S., J. F. Scott, and A. Kornberg. 1976. An enzyme system for replication of duplex circular DNA: the replicative form of phage ϕX174. Proc. Natl. Acad. Sci. USA 73:1594–1597.
42. Ekechukwu, M. C., and B. A. Fane. 1995. Characterization of the morphogenetic defects conferred by cold-sensitive prohead accessory and scaffolding proteins of ϕX174. J. Bacteriol. 177:829–830.
43. Ekechukwu, M. C., D. J. Oberste, and B. A. Fane. 1995. Host and ϕX174 mutations affecting the morphogenesis or stabilization of the 50S complex, a single-stranded DNA synthesizing intermediate. Genetics 140:1167–1174.
44. Fane, B. A. Unpublished results.
45. Fane, B. A., and M. Hayashi. 1991. Second-site suppressors of a cold-sensitive prohead accessory protein of bacteriophage ϕX174. Genetics 128:663–671.
46. Fane, B. A., S. Shien, and M. Hayashi. 1993. Second-site suppressors of a cold sensitive external scaffolding protein of bacteriophage ϕX174. Genetics 134 1003–1011.
47. Feige, U. and S. Stirm. 1976. On the structure of the *Escherichia coli* C cell wall lipopolysaccharide core and its ϕX174 receptor region. Biochem. Biophys. Res. Commun. 71:566–573.
48. Floor, E. 1970. Interaction of morphogenetic genes of bacteriophage T4. J. Mol. Biol. 47:293–306.
49. Fluit, A. C., P. D. Baas, and H. S. Jansz. 1985. The complete 30-base pair origin region of bacteriophage ϕX174 in a plasmid is both required and sufficient for in vivo rolling circle DNA replication. Eur. J. Biochem. 149:579–584.

50. Fujisawa, H., and M. Hayashi. 1976. Viral DNA-synthesizing intermediate complex isolated during assembly of bacteriophage ϕX174. J. Virol. 19:409–415.
51. Georgopoulos, C. P., and B. Hohn. 1978. Identification of a host protein necessary for bacteriophage morphogenesis (the *groE* gene product). Proc. Natl. Acad. Sci. USA 75:131–135.
52. Godson, G. N., B. G. Barrell, R. Standen, and F. C. Fiddes. 1978. Nucleotide sequence of bacteriophage G4 DNA. Nature 276:236–247.
53. Gordon C. L., S. K. Sather, S. Casjens, and J. King. 1994. Selective in vivo rescue by GroEL/ES of thermolabile folding intermediates to phage P22 structural proteins. J. Biol. Chem. 269:27941–2751.
54. Haanes, E. J., D. R. Thomsen, S. Martin, F. L. Homa, and D. E. Lowery. 1995. The bovine herpesvirus 1 maturational proteinase and scaffold proteins can substitute for the homologous herpes simplex virus type 1 proteins in the formation of hybrid type B capsids. J. Virol. 69:7375–7379.
55. Hafenstein, S., and B. A. Fane, unpublished results.
56. Hafenstein, S., and B. A. Fane. 2002. ϕX174 genome–capsid interactions influence the biophysical properties of the virion: evidence for a scaffolding-like function for the genome during the final stages of morphogenesis. J. Virol. 76:5350–5356.
57. Hamatake, R. K., A. Aoyama, and M. Hayashi. 1985. The *J* gene of ϕX174: In vitro analysis of J protein function. J. Virol. 54:345–350.
58. Hamatake, R. K., K. J. Buckley, and M. Hayashi. 1988. The *J* gene of ϕX174: isolation and characterization of a J gene mutant. Mol. Gen. Genet. 211:72–77.
59. Hanninen, A. L., D. Bamford, and J. K. Bamford. 1997. Assembly of membrane-containing bacteriophage PRD1 is dependent on GroEL and GroES. Virology 227:207–210.
60. Hayashi, M. Personal communication.
61. Hayashi, M. 1978. Morphogenesis of the isometric phages, pp. 531–547. *In* D. Denhardt, D. Dressler, and D. S. Ray (eds.) The Single Stranded DNA Phages. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
62. Hayashi, M., A. Aoyama, D. L. Richardson, and M. N. Hayashi. 1988. Biology of the bacteriophage ϕX174, pp. 1–71. *In* R. Calendar (ed.) The Bacteriophages, vol. 2. Plenum Press, New York.
63. Hayashi, M., F. K. Fujimura, and M. Hayashi. 1976. Mapping of in vivo messenger RNAs for bacteriophage ϕX174. Proc. Natl. Acad. Sci. USA 73:3519–3523.
64. Hayashi, M. N., and M. Hayashi. 1981. Stability of bacteriophage ϕX174-specific mRNA in vivo. J. Virol. 37:506–510.
65. Hayashi, M. N., and M. Hayashi. 1985. Cloned DNA sequences that determine mRNA stability of bacteriophage ϕX174 in vivo are functional. Nucleic Acids Res. 13:5937–5948.
66. Hendrix, R. W., L. G. Lawrence, G. F. Hatfull, and S. Casjens. 2000. The origins and ongoing evolution of viruses. Trends Microbiol. 8:504–508.
67. Hendrix, R. W., M. C. Smith, N. Burns, E. F. Ford, and G. F. Hatfull. 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. Proc. Natl. Acad. Sci. USA 96:2192–2197.
68. Hendrix R. W., and L. Tsui. 1978. Role of the host in virus assembly: cloning of the *Escherichia coli* groE gene and identification of its protein product. Proc. Natl. Acad. Sci. USA 75:136–139.
69. Holder, K. K., and J. J. Bull. 2001. Profiles of adaptation in two similar viruses. Genetics 159:1393–1404.
70. Hutchison, C.A. III, and R. L. Sinsheimer. 1966. The process of infection with bacteriophage phi-X174. X. Mutations in a ϕX lysis gene. J. Mol. Biol. 18:429–447.
71. Ikeda, M., M. Wachi, H. K. Jung, F. Ishino, and M. Matsuhashi. 1991. The *Escherichia coli mraY* gene encoding UDP-*N*-acetylmuramoyl-pentapeptide: undecaprenyl-phosphate phospho-*N*-acetylmuramoyl-pentapeptide transferase. J. Bacteriol. 1173:1021–1026.
72. Ilag, L. L., N. H. Olson, T. Dokland, C. L. Music, R. H. Cheng, Z. Brown, R. McKenna, M. G. Rossmann, T. S. Baker, and N. L. Incardona. 1995. Bacteriophage ϕX174 procapsid: purification and structure at 25 Å resolution. Structure 3:353–363.
73. Ilag, L. L., R. McKenna, M. P. Yadav, J. N. BeMiller, N. L. Incardona, and M. G. Rossmann. 1994. Calcium ion-induced structural changes in bacteriophage ϕX174. J. Mol. Biol. 244:291–300.
74. Ilag, L. L., J. K. Tuech, L. A. Beisner, R. A. Sumrada, and N. L. Incardona. 1993. Role of DNA–protein interactions in bacteriophage phi X174 DNA injection. J. Mol. Biol. 229:671–684.
75. Incardona, N. L. 1983. A kinetic model for virus binding which involves release of cell-bound virus–receptor complexes. J. Theor. Biol. 105:631–645.
76. Incardona, N.L., and L. Selvidge. 1973. Mechanism of adsorption and eclipse of bacteriophage phi X174. II. Attachment and eclipse with isolated *Escherichia coli* cell wall lipopolysaccharide. J. Virol. 11:775–782.
77. Incardona, N. L., J. K. Tuech, and G. Murti. 1985 Irreversible binding of phage ϕX174 to cell-bound lipopolysaccharide receptors and release of virus–receptor complexes. Biochemistry 24:6439–6446.
78. Inagaki, M., A. Tanaka, R. Suzuki, H. Wakashima, T. Kawaura, S. Karita, S. Nishikawa, and N. Kashimura. 2000. Characterization of the binding of spike H protein of bacteriophage phiX174 with receptor lipopolysaccharides. J. Biochem. (Tokyo) 127:577–583.
79. Jazwinski, S. M., A. A. Lindberg, and A. Kornberg. 1975. The gene H spike protein of bacteriophages ϕX174 and S13. I. Functions in phage-receptor recognition and in transfection. Virology 66:283–293.
80. Jazwinski, S. M., A. A. Lindberg, and A. Kornberg. 1975. The lipopolysaccharide receptor for bacteriophage phiX174 and S13. Virology 66:268–282.
81. Jennings, B., and B. A. Fane. 1997. Genetic analysis of the ϕX174 DNA binding protein. Virology 227:370–377.
82. Kawaura, T., M. Inagaki, S. Karita, M. Kato, S. Nishikawa, and N. Kashimura. 2000. Recognition of receptor lipopolysaccharides by spike G protein of bacteriophage ϕX174. Biosci. Biotechnol. Biochem. 64:1993–1997.
83. Kodaira, K., K. Nakano, S. Okada, and A. Taketo. 1992. Nucleotide sequence of the genome of bacteriophage α3: interrelationship of the genome structure and the gene

products with those of the phages ϕX174, G4 and ϕK. Biochim. Biophys. Acta 1130:277–288.

84. Kornberg, A. 1980. DNA Replication. Freeman, San Francisco.
85. Kornberg, A. 1982. Supplement to DNA Replication. Freeman, San Francisco.
86. Krol, M. A., N. H. Olson, J. Tate, J. E. Johnson, T. S. Baker, and P. Ahlquist. 1999. RNA-controlled polymorphism in the in vivo assembly of 180-subunit and 120-subunit virions from a single capsid protein. Proc. Natl. Acad. Sci. USA 96:13650–13655.
87. Leffers, G., and V. B. Rao. 1996. A discontinuous headful packaging model for packaging less than headful length DNA molecules by bacteriophage T4. J. Mol. Biol. 258:839–850.
88. Liu, B. L., J. S. Everson, B. A. Fane, P. Giannikopoulou, E. Vretou, P. R. Lambden, and I. N. Clarke. 2000. The molecular characterization of a bacteriophage (Chp2) from *Chlamydia psittaci*. J. Virol. 74:3646–3649.
89. Low, R. L., J. Shlomai, and A. Kornberg. 1982. Protein n, a primosomal DNA replication protein of *Escherichia coli*. Purification and characterization. J. Biol. Chem. 257:6242–6250.
90. Maratea, D., K. Young, and R. Young. 1985. Deletion and fusion analysis of the phage ϕX174 lysis gene E. Gene 40:39–46.
91. McKenna, R. Personal communication.
92. McKenna, R., B. R. Bowen, L. L. Ilag, M. G. Rossmann, and B. A. Fane. 1996. The atomic structure of the degraded procapsid particle of bacteriophage G4: induced structural changes in the presence of calcium ions and functional implications. J. Mol. Biol. 265:736–750.
93. McKenna, R., L. L. Ilag, and M. G. Rossmann. 1994. Analysis of the single-stranded DNA bacteriophage ϕX174 at a resolution of 3.0 Å. J. Mol. Biol. 237:517–543.
94. McKenna, R., D. Xia, P. Willingmann, L. L. Ilag, S. Krishnaswamy, M. G. Rossmann, N. H. Olson, T. S. Baker, and N. L. Incardonna. 1992. Atomic structure of single-stranded DNA bacteriophage ϕX174 and its functional implications. Nature 355:137–143.
95. Mukai, R., R. K. Hamatake, and M. Hayashi. 1979. Isolation of the bacteriophage ϕX174 prohead. Proc. Natl. Acad. Sci. USA 76:4877–4881.
96. Munekiyo, R., T. Tsuzuki, and M. Sekiguchi. 1979. A new locus of *Escherichia coli* that determines sensitivity to bacteriophage ϕX174. J. Bacteriol. 138:1038–1040.
97. Nakonechny, W. S., and C. M. Teschke. 1998. GroEL and GroES control of substrate flux in the in vivo folding pathway of phage P22 coat protein. J. Biol. Chem. 273:27236–27244.
98. Newbold, J. E., and R. L. Sinsheimer. 1970. The process of infection with bacteriophage ϕX174. XXXII. Early steps in the infection process: attachment, eclipse and DNA penetration. J. Mol. Biol. 49:49–66.
99. Newcomb, W. W., B. L. Trus, N. Cheng, A. C. Steven, A. K. Shaeffer, D. J. Tenney, S. K. Weller, and J. C. Brown. 2000. Isolation of herpes simplex virus procapsids from cells infected with a protease-deficient mutant virus. J. Virol. 74:1663–1673.
100. Novak, C., and B. A. Fane. 2004. The functions of the N-terminus of the ϕX174 internal scaffolding protein, a protein encoded in an overlapping reading frame in a two scaffolding protein system. J. Mol. Biol. 335:383–390.
101. Oberste, J., and B. A. Fane. Unpublished results.
102. Prasad, B. V. V., P. E. Prevelige, E. Marietta, R. O. Chen, D. Thomas, J. King, and W. Chui. 1993. Three-dimensional transformation of capsids associated with genome packaging in a bacterial virus. J. Mol. Biol. 231:65–74.
103. Prevelige, P. E. Jr., D. Thomas, and J. King. 1993. Nucleation and growth phases in the polymerization of coat and scaffolding subunits into icosahedral procapsid shells. Biophys. J. 64:824–835.
104. Renaudin, J., M. C. Paracel, and J. M. Bove. 1987. Spiroplasma virus 4: nucleotide sequence of the viral DNA, regulatory signals and the proposed genome organization. J. Bacteriol. 169:4950–4961.
105. Richardson, D. L. Jr., A. Aoyama, and M. Hayashi. 1988. Proteolysis of bacteriophage ϕX174 prohead protein gpB by a protease located in the *Escherichia coli* outer membrane. J. Bacteriol. 170:5564–5571.
106. Roof, W. D., S. M. Horne, K. D. Young, and R. Young. 1994. SlyD, a host gene required for ϕX174 lysis, is related to the FK506-binding protein family of peptidyl-prolyl cis-trans-isomerases. J. Biol. Chem. 269:2902–2910.
107. Sanger, F., A. R. Coulson, C. T. Friedmann, G. M. Air, B. G. Barrell, N. L. Brown, J. C. Fiddes, C. A. Hutchison III, P. M. Slocombe, and M. Smith. 1978. The nucleotide sequence of bacteriophage ϕX174. J. Mol. Biol. 125:225–246.
108. Savithri, H. S., and J. W. Erickson. 1983. The self-assembly of the cowpea strain of southern bean mosaic virus: formation of $T = 1$ and $T = 3$ nucleoprotein particles. Virology 126:328–335.
109. Schmidt, G. 1973. Genetical studies on the lipopolysaccharide structure of *Escherichia coli* K12. J. Gen. Microbiol. 77:151–160.
110. Scott, J. F., S. Eisenberg, L. L. Bertsch, and A. Kornberg. 1977. A mechanism of duplex DNA replication revealed by enzymatic studies of phage ϕX174: catalytic strand separation in advance of replication. Proc. Natl. Acad. Sci. USA 74:193–197.
111. Sertic, V., and N. Bulgakov. 1935. Classification et identification des typhi-phage. C. R. Soc. Biol. Paris. 119:1270–1272.
112. Serwer, P., and M. E. Pichler. 1978. Electrophoresis of bacteriophage T7 and T7 capsids in agarose gels. J. Virol. 28:917–928.
113. Shlomai, J., and A. Kornberg. 1980. An *Escherichia coli* replication protein that recognizes a unique sequence within a hairpin region in ϕX174 DNA. Proc. Natl. Acad. Sci. USA 77:799–803.
114. Shlomai, J., L. Polder, K. Arai, and A. Kornberg. 1981. Replication of ϕX174 DNA with purified enzymes. I. Conversion of viral DNA to a supercoiled, biologically active duplex. J. Biol. Chem. 256:5233–5238.
115. Siden, E. J. and M. Hayashi. 1974. Role of the gene B product in bacteriophage ϕX174 development. J. Mol. Biol. 89:1–16.
116. Simpson, A. A., Y. Tao, P. G. Leiman, M. O. Badasso, Y. He, P. J. Jardine, N. H. Olson, M. C. Morais, S. Grimes,

D. L. Anderson, T. S. Baker, and M. G. Rossmann. 2000. Structure of the bacteriophage ϕ29 DNA packaging motor. Nature 408:745–750.

117. Sinsheimer, R. L. 1959. A single-stranded deoxyribonucleic acid from bacteriophage ϕX174. J. Mol. Biol. 1:43–53.
118. Sinsheimer, R. L. 1968. Bacteriophage ϕX174 and related viruses. Prog. Nucleic Acid Res. Mol. Biol. 8:115–169.
119. Smith, L. H., K. Grohmann, and R. L. Sinsheimer. 1974. Nucleotide sequences of the 5′ termini of ϕX174 mRNAs synthesized in vitro. Nucleic Acids Res. 1:1521–1529.
120. Smith, L. H., and R. L. Sinsheimer. 1976. The in vitro transcription units of bacteriophage ϕX174. I. Characterization of synthetic parameters and measurement of transcript molecular weights. J. Mol. Biol. 103:681–697.
121. Smith, L. H., and R. L. Sinsheimer. 1976. The in vitro transcription units of bacteriophage ϕX174. II. In vitro initiation sites of ϕX174 transcription. J. Mol. Biol. 103:699–710.
122. Smith, L. H., and R. L. Sinsheimer. 1976. The in vitro transcription units of bacteriophage ϕX174. III. Initiation with specific 5′ end oligonucleotides of in vitro ϕX174 RNA. J. Mol. Biol. 103:711–735.
123. Spindler, K.R., and M. Hayashi. 1979. DNA synthesis in *Escherichia coli* cells infected with gene H mutants of bacteriophage ϕX174. J. Virol. 29:973–982.
124. Sternberg, N. 1973. Properties of a mutant of *Escherichia coli* defective in bacteriophage lambda head formation (groE). I. Initial characterization. J. Mol. Biol. 76:1–23.
125. Sternberg, N. 1976 A genetic analysis of bacteriophage lambda head assembly. Virology 71:568–582.
126. Storey, C. C., M. Lusher, and S. J. Richmond. 1989. Analysis of the complete nucleotide sequence of Chp1, a phage which infects *Chlamydia psittaci*. J. Gen. Virol. 70:3381–3390.
127. Sullivan, W. 1979. New York Times, 7 May, p. D13.
128. Tessman, E. S., and P. K. Peterson. 1976. Bacterial *rep*$^{-}$ mutations that block development of small DNA bacteriophages late in infection. J. Virol. 20:400–412.
129. Tessman, E. S., I. Tessman, and T. J. Pollock. 1980. Gene K of bacteriophage ϕX174 codes for a nonessential protein. J. Virol. 33:557–560.
130. Tetart, F., C. Desplats, M. Kutatrladze, C. Monod, H.-W. Ackermann, and H. M. Krisch. 2001. Phylogeny of the major head and tail genes of the wide-ranging T4-type bacteriophages. J. Bacteriol. 183:358–366.
131. Thuman-Commike, P. A., B. Greene, J. A. Malinski, M. Burbea, A. McGough, W. Chiu, and P. E. Prevelige Jr. 1999. Mechanism of scaffolding-directed virus assembly suggested by comparison of scaffolding-containing and scaffolding-lacking P22 procapsids. Biophys. J. 76:3267–3277.
132. Tonegawa, S. and M. Hayashi. 1970. Intermediates in the assembly of ϕX174. J. Mol. Biol. 48:19–42.
133. Valentine, C. R., B. A. Montgomery, S. G. Miller, R. R. Delongchamp, B. A. Fane, and H. V. Malling. 2002. Characterization of mutant spectra generated by a forward mutational assay for gene A of ϕX174 from enu-treated transgenic mouse embryonic cell line PX-2. Environ. Mol. Mut. 39:55–68.
134. Van Masnfeld, A. D., P. D. Baas, and H. S. Jansz. 1984. Gene A protein of bacteriophage ϕX174 is a highly specific single-stranded DNA nuclease and binds via a tyrosyl residue to DNA after cleavage. Adv. Exp. Med. Biol. 197:221–230.
135. Van Mansfeld, A. D., S. A. Langeveld, P. J. Weisbeek, P. D. Baas, G. A. van Arkel, and H. S. Jansz. 1979. Cleavage site of ϕX174 gene-A protein in ϕX and G4 RFI DNA. Cold Spring Harb. Symp. Quant. Biol. 43:331–334.
136. Weisbeek, P. J., J. H. van de Pol, and G. A. van Arkel. 1983. Mapping of host range mutants of bacteriophage ϕX174. Virology 52:408–416.
137. Wichman, H. A., L. A. Scott, C. D. Yarber, and J. J. Bull. 2000. Experimental evolution recapitulates natural evolution. Phil. Trans. R. Soc. Lond. B. Biol. Sci. 355:1677–1684.
138. Yazaki, K. 1981. Electron microscopic studies of bacteriophage ϕX174 intact and "eclipsing" particles, and the genome by the staining and shadowing method. J. Virol. Methods 2:159–167.
139. Young, K.D., and R. Young. 1982. Lytic action of cloned ϕX174 gene E. J. Virol. 44:993–1002.
140. Young, R. 1992. Bacteriophage lysis: mechanism and regulation. Microbiol. Rev. 56:430–81.
141. Zhou, Z. H., J. S. Macnab, J. Jakana, L. R. Scott, W. Chiu, and F. J. Rixon. 1998. Identification of the sites of interaction between the scaffold and outer shell in HSV-1 capsids by difference electron imaging. Proc. Natl. Acad. Sci. USA 95:2778–2783.

12

Filamentous Phage

MARJORIE RUSSEL
PETER MODEL

Filamentous phages constitute a large family of bacterial viruses that infect many Gram-negative bacteria and even a Gram-positive bacterium (18). These long, slender viruses contain a circular, single-stranded DNA genome encased in a somewhat flexible tube composed of thousands of copies of a single major coat protein (figure 12-1). Two minor proteins at one end and two others at the other end seal the tips of the tube. The genome consists of a dozen or fewer closely packed genes and an intergenic (IG) region that contains sequences necessary for DNA replication and encapsidation. Unlike most bacterial viruses, filamentous phage are produced and secreted from infected bacteria without cell killing or lysis. Rather, they assemble at and are secreted across the cell membrane(s). Readers are referred to several other reviews on filamentous phage (4, 82, 101, 106, 112) for more comprehensive information and citations of the primary literature than are given here.

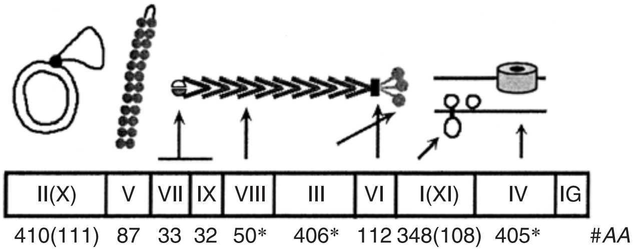


Figure 12-1 Filamentous phage f1/M13/fd: genes and gene products. pII binds to a sequence (the + strand origin) in the intergenic region (IG) of double-stranded DNA and nicks the (+) strand; the original (+) strand is displaced by Rep helicase as a new (+) strand is elongated from the 3′ end of the nick by host DNA polymerase III, using the (−) strand as template. pX, which is identical to the C-terminal third of pII, is required for the accumulation of single-stranded DNA, as is pV. Dimers of pV bind cooperatively to single-stranded DNA, which collapses the circular genome into a flexible rod with the packaging signal (PS) exposed at one end of the filament. pVII and pIX are small coat proteins located at the tip of the virus that is first to emerge from the cell during assembly. pVIII is the major coat protein, several thousand copies of which form the cylinder that encases the single-stranded DNA phage genome. pIII and pVI are located at the end of the virion where they mediate termination of assembly and release of the virion from the cell membrane. pIII is also necessary for phage infectivity. pI may hydrolyze ATP to promote assembly; pXI is identical to the C-terminal third of pI; it lacks the cytoplasmic domain and may play a structural role as part of an oligomeric pI/pXI complex. pIV is a multimeric outer membrane channel through which the phage exits the bacterium.

Most information about filamentous phages derives from those that infect *Eschericia coli* (f1/M13/fd, and to a lesser extent Ike and I2-2). Structural analysis of several others has shown that the packing density (protein:DNA mass ratio in the particle) can vary, as can the symmetry of the particles. Those defined as class I, including the *E. coli* phages, have a 5-fold rotation axis combined with a 2-fold screw axis, while class II phage like *Pseudomonas* phages Pf1 and Pf3 have a simple one-start helix (107). The mechanism of assembly is likely to be fundamentally the same for both classes; perhaps differences in the primary sequences of the major coat proteins account for the different subunit packing. DNA sequences reveal the modular nature of phage evolution and other variations. Phages I2-2 and IKe are highly homologous over the two thirds of their genomes that encode capsid and morphogenetic proteins, whereas the remaining portions, containing the replication origins and replication genes, are unrelated (153). Although Pf3 has a gene IV, it is in a different position in the genome than in other filamentous phages.*Vibrio cholerae* CTXΦ lacks gene IV (162, 163), which encodes an outer membrane assembly protein that is essential in the *E. coli* phages; instead, it uses the homologous host protein, EpsD (26). EpsD, one of three *Vibrio* pIV homologs (members of the widespread “secretin” family), is also required for secretion of cholera toxin (133, 143). B5, the recently discovered filamentous phage that infects the Gram-positive *Propionibacterium freudenreichii*, also lacks a gene IV (18); since Gram-positive bacteria lack an outer membrane, presumably neither pIV nor a bacterial equivalent is necessary. Whereas adjacent genes encode two small minor coat proteins (pVII and pIX) located at the same end of the particle in *E. coli* phages, a single gene is located in the analogous region of the Pf3 and some other phage genomes, which may encode a functionally equivalent fused minor coat protein. The major coat proteins of the *E. coli* phages are synthesized with a signal sequence, while those of Pf3 (90), PH75, a filamentous phage of *Thermus thermophilus* (119), and B5 (18) are not. Several filamentous phages of *Vibrio* and *Xanthomonas*, but not those of *E. coli* or *Pseudomonas aeruginosa*, lysogenize their host (10, 93, 163). Finally, CTXΦ is so far unique among filamentous phages (although not among phages in general) in carrying additional genes, in this case the genes that encode cholera toxin (163). Unless specified, the properties of filamentous phage described below refer to the almost identical f1, M13, and fd (the "Ff," for F-specific filamentous phage, see "Infection" below).

Structure of the Phage Particle

The image of a filamentous phage obtained by atomic force microscopy is shown in figure 12-2. They are about 6.5 nm in diameter, with a length determined by the size of the genome, normally 6–7 kb of single-stranded DNA. The 6400-nucleotide Ff genome is encapsidated in a 930 nm particle; a 221-nucleotide "microphage" variant is only 50 nm long (152), and the presence of cloned DNA in the phage genome makes the particle proportionately longer. Although there is no theoretical limit to the amount of DNA that can be packaged in a single particle, as there is for icosahedral phages of fixed dimensions, there may be a practical limit; phage containing about 6 kb of additional DNA make small plaques, so even larger inserts would likely have even more deleterious effects. Even normal particle length is selected against when phage genes are made dispensable (by providing them from another source)—smaller particles accumulate with genomes that contain *cis*-active replication and packaging signals but no intact genes (71). Much longer particles—10 or more times the normal unit length—can be generated by eliminating pIII, the phage protein necessary to terminate assembly and release the particle (and to mediate infection), but these are noninfectious particles that contain multiple unit-length genomes (122).

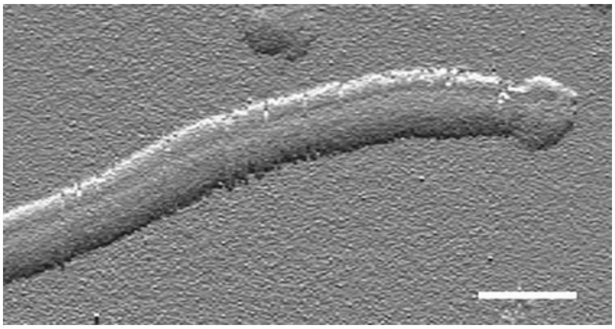


Figure 12-2 Filamentous phage visualized by atomic force microscopy. The large size of the phage tip structure suggests that it is the pIII–pVI end. The lollipop-like images of pIII sometimes observed by electron microscopy may represent partially denatured pIII.

The protein tube that surrounds the single-stranded DNA is composed of several thousand copies of pVIII, the 50-residue major coat protein (in the Ff), oriented at a 20° angle from the particle axis and overlapped like fish scales to form a right-handed helix (106). The filament is held together by interactions between the hydrophobic midsections of adjacent subunits; other crucial interactions have been postulated between the hydrophobic face of an N-terminal amphipathic helix on one subunit with the hydrophobic face of a C-terminal amphipathic helix on a neighboring subunit (129). Except for five surface-exposed N-terminal residues, each pVIII subunit forms a single, continuous α-helix. The positively charged residues near the C-terminus are at the inner surface of the tube and interact with phosphates of the viral single-stranded DNA (54, 73, 130).

Electron micrographs reveal that each end of the particle has a distinct morphology. The blunt end contains three to five copies each of pVII and pIX, two of the smallest ribosomally translated proteins with a defined function that are known (33 and 32 residues respectively). Immunological evidence indicates that at least some of pIX is exposed (36), and since N-terminal display is possible on both (48), both N-termini are probably at or near the surface. Both proteins are quite hydrophobic; the N-terminal portions of pVII from several phages are negatively charged, whereas there are two or three positively charged residues near the C-terminus of the pIXs. Phage assembly begins at the pVII–pIX end, and if either protein is absent, particle production does not occur (98).

The other end of the particle is pointed, and lollipop-like knobs can be seen to extend from the tip in certain preparations (52). This end contains about five copies each of pIII and pVI, the proteins that mediate phage entry and exit. The N-terminal domain of pIII binds the infecting phage to its cellular receptors, the tip of the F pilus and the C-terminal domain of TolA (11, 19), and pVI must be present in order for pIII to be incorporated into the particle (121). These two proteins are also necessary for phage assembly to terminate, that is for newly formed phage to detach from the cell membrane. If either is absent, "polyphage," which contain multiple phage genomes encased in particles 10–20 times the normal phage length, accumulate and remain tethered to the cell (122). A complex composed of pIII and pVI can be isolated from phage particles (47). Although they probably do associate in the cell membrane (since pVI is degraded in cells that lack pIII), no complex could be detected, which is consistent with the proposal that detachment involves a conformational change in the complex that increases its stability (121). The disposition of pVI in the particle is not known; pVI with fusions at the C-terminus can be incorporated into phage (albeit inefficiently), suggesting that this portion of the 112-residue pVI can be exposed (121). The N-terminal domain of the 406-residue pIII is surface exposed and is responsible for the "knob" structures (2). The C-terminal 132 residues of pIII are necessary and sufficient for incorporation into the phage particle, termination of assembly, and release of phage from the cell; this domain is likely to be buried within the particle (121).

The single-stranded phage genome is oriented and anchored within the phage particle by the packaging signal (PS), which is located in the noncoding IG region of the genome and is positioned at the pVII–pIX end of the particle (56, 168). The PS, an imperfect but extremely stable hairpin, is necessary and sufficient for efficient encapsidation of circular single-stranded DNA into phage particles, whereas an "artificial" PS (a perfect duplex of equivalent length) is a poor substitute (140). Certain amino acid substitutions in pVII, pIX, and pI (see below) enable single strands that lack a PS to be encapsidated; it is not known whether the DNA is randomly oriented in such particles or whether some small duplex region serves as a secondary PS.

Infection

A simplified cartoon illustrating the filamentous phage life cycle is shown in figure 12-3. Filamentous phages use pili as primary receptors to infect cells. Pili are long filamentous structures that emanate from the cell surface. There are different types of pili; those utilized by the phage are anchored in the cytoplasmic membrane and are capable of retraction—the subunits depolymerize back into the membrane. The *E. coli* phages use conjugative pili, self-transmissible pili that mediate transfer of the plasmid that encodes them to recipient bacteria (147). The f1/M13/fd phages (which are almost identical in sequence and are referred to collectively as Ff) bind F pili, while IKe uses N or P pili. CTXΦ uses nonconjugative type IV pili (Tfp) (163). Tfp mediate adhesion to eukaryotic cells and are responsible for a form of solid-surface locomotion called "twitching motility" (164). Phage can infect cells that lack appropriate pili, but the process is extremely inefficient; the efficiency is improved 2–4 orders of magnitude by agents that concentrate the phage or promote its adherence to the cell surface, such as calcium chloride and polyethylene glycol (142).

Two surface-exposed N-terminal domains of the minor coat protein pIII mediate infection by binding first to the tip of the F pilus and then to the host TolA protein (155). In the Ff pIII, the N1 (or D1) domain (residues 1–68) that binds TolA precedes the pilus-binding domain (the N2 or D2 domain, residues 87–217) (28), but the position of the pilus-binding and TolA-binding domains is reversed in IKe (34). The Ff and IKe pIII proteins are poorly conserved and cannot replace one another to assemble into the heterologous phage (15, 35). Ff pIII has a second glycine-rich domain after N2 that causes cells to leak periplasmic contents and become hypersensitive to detergents, whereas IKe pIII has

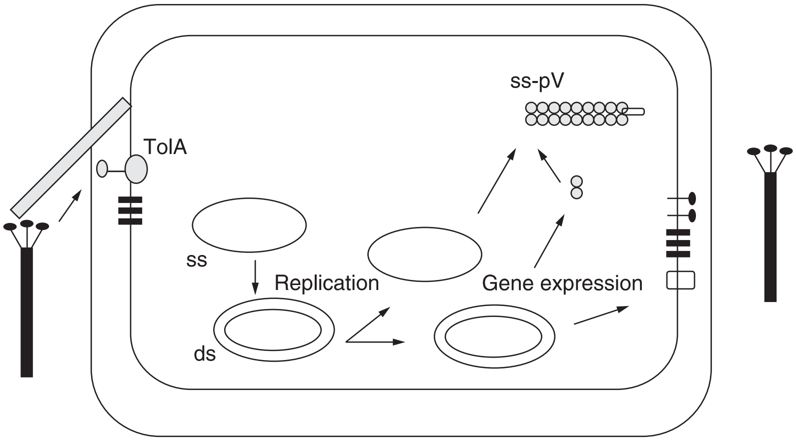


Figure 12-3 Life cycle of filamentous phage f1/M13/fd. The phage (via pIII, located at one end) binds to the tip of the F pilus. Pilus retraction brings the phage close to the cell surface where pIII then binds to the periplasmic domain of host TolA. The cytoplasmic membrane-anchored TolA, Q, and R proteins mediate depolymerization of the phage coat proteins into the membrane (where they are available for reutilization) and entry of the phage circular single-stranded DNA into the cytoplasm. The single-stranded DNA is converted to a double-stranded, supercoiled replicative form by the action of host RNA polymerase, DNA polymerase III, and gyrase. Supercoiled RF is a template for phage gene expression and rolling circle replication, which generates a single-stranded DNA molecule. When sufficient pV has accumulated, pV dimers cover the single-stranded DNA, leaving the PS hairpin exposed at one end; the pV–single-stranded DNA complex is a substrate for assembly. The five coat proteins are integral cytoplasmic membrane proteins prior to their incorporation into the phage particle.

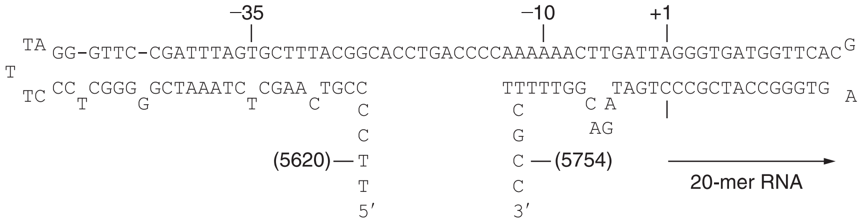

−35
−10
+1
TA GG-GTTC-CGATTTAGTGCTTTACGGCACCTGACCCCAAAAAACTTGATTAGGGTGATGGTTCAC G
T
TC CC CGGG GCTAAATC CGAA TGCC
T G T C
C
C
(5620)— T
T
5′
TTTTTGG TAGTCCCGCTACCGGGTG A
C A
AG
C
G
C —(5754)
C
3′
20-mer RNA

Figure 12-4 The (−) strand origin of f1 functions as an unusual promoter. The structure of the (−) strand origin as deduced from footprinting and mutational studies. Adapted from (62).

a single glycine-rich domain located between N1 and N2 and does not induce outer membrane defects (9, 15, 123).

Pili normally assemble and disassemble continuously, and this, possibly stimulated by phage binding, brings the phage close to the cell surface. Crystallographic analyses indicate that the N2 and N1 domains of pIII interact as a horseshoe-shaped structure (66, 67, 100). Binding of the N2 domain to the pilus releases the N1 domain, making it available to bind to the C-terminal domain of the host TolA protein, TolA-III (99, 126). Mutational studies implicate the outer surface of the N2 portion of the "horseshoe" in F-pilus binding (27). TolA-III, which may contact the outer membrane, is tethered via a long periplasmic linker (TolA-II) to TolA-I, which spans the cytoplasmic membrane and interacts with the membrane proteins TolQ and TolR (19, 167). These three Tol proteins are absolutely required for phage infection (158, 159). How the phage penetrates the outer membrane and the underlying peptidoglycan layer to engage TolA is not known. The Tol proteins mediate depolymerization of the phage coat proteins into the cytoplasmic membrane and translocation of the viral single-stranded DNA into the bacterial cytoplasm (20). The molecular details of how this is accomplished are completely obscure, though it has been suggested that the conversion of phage filaments into spheroid-shaped particles at chloroform–water interfaces may mimic a spontaneous event that occurs when the phage meets a membrane (30, 56). The precise role of the Tol proteins with respect to the bacterium is not known, but they are important for maintaining the integrity of the bacterial outer membrane; bacteria that lack a Tol protein are hypersensitive to detergents and leak their periplasmic contents, which mimics the effect of the presence of Ff pIII. Recently a role for TolA in delivering lipopolysaccharides to the outer membrane has been demonstrated (49), and it has been proposed that the Tol proteins, which are analogs of the TonB system, may transmit energy from the proton-motive force to the outer membrane (96).

Replication

Once the viral single-stranded phage DNA (the (+) strand) enters the cytoplasm, host enzymes convert it to a double-stranded, supercoiled molecule (RF or replicative form). An RNA primer is required to synthesize the (−) strand. This 20 nt primer is generated by RNA polymerase, which initiates synthesis at an unusual site (the (−) strand origin, figure 12-4) located in the noncoding intergenic (IG) region of the (+) strand (63, 64, 72). The site consists of two adjacent hairpins separated by a single-stranded region; one hairpin sequence includes a promoter-like −35 motif and the other has a –10 motif (72); the affinity of RNA polymerase to this structure is much greater than to an authentic promoter such as the lacUV5 promoter (62). The RNA primer is extended by host DNA polymerase III, the newly replicated strand is ligated, and the double-stranded product is supercoiled by gyrase. RF is the template for phage gene expression. Phage gene expression—in particular, synthesis of pII, a site-specific nicking-closing enzyme—is necessary for further replication of the initial RF. pII nicks the (+) strand of the RF at the (+) strand origin in the IG region (53, 72). This is a complex site that includes sequences for binding IHF, pII, nicking, and a site necessary for termination of synthesis. The 3′ end of the nick is elongated by host DNA polymerase III using the (−) strand as template (61). The original (+) strand is displaced by Rep helicase as the new (+) strand is synthesized, and when a round of replication is complete the displaced (+) strand is recircularized by the nicking-closing activity of pII and again converted to RF.

Early after infection when the concentration of the phage-encoded single-stranded DNA-binding protein (pV) is low, newly synthesized single strands are immediately converted to RF, and both RF and phage proteins increase exponentially. As its concentration increases, pV binds cooperatively to newly generated (+) strands. This covers them and prevents polymerase access, thereby blocking their conversion to RF. pX, which is identical to the C-terminal 111 residues of pII, is required for the stable accumulation of single strands at this stage, but the mechanism by which it acts is not known (45, 46). pV is dimeric, with the interaction surface of the subunits opposite the DNA-binding surface (88, 148). Upon binding, the back-to-back arrangement of the dimers collapses the circular single strand into a rodlike structure (51). The DNA is oriented in this complex, with the PS hairpin protruding from one end (5). This presumably occurs because pV binds double-stranded DNA only weakly. The fact that Y-shaped or other more complicated pV–ssDNA structures (which would indicate multiple initial binding events) are not seen in vivo suggests that there may be a favored nucleation point for pV binding. pV does bind preferentially to the DNA analog of the site on gene II mRNA (108) at which it represses translation (44, 109), but this site is 400 nucleotides away from the PS; ideally, a nucleation site should be adjacent to the PS so that cooperative interactions between pV dimers would "zip" the opposing portions of the single-stranded circle into the rod.

Genes and Gene Expression

The Ff genome (~6400 nts) contains nine closely packed genes and one major noncoding region (the IG) which contains the (+) and (−) strand replication origins and the PS (figure 12-1). Two of the phage genes encode two proteins, for a total of 11 phage-encoded proteins. Genes I (124) and II (45) have internal translational initiation sites from which in-frame restart proteins are produced. In each case, both the full-length and the restart protein (whose sequence is identical to the C-terminal third of the full-length protein) are necessary for successful phage production. The relative levels of pII and pX determine the distribution of phage DNA species in infected hosts (87). The absence of pII totally prevents phage DNA synthesis, while the absence of its restart partner, pX, prevents accumulation of (+) strands (46). The pI/pXI pair is described below. Of the 11 phage-encoded proteins, three (pII, pX, pV) are required to generate single stranded DNA, three (pI, pXI, pIV) are required for phage morphogenesis, and five (pIII, pVI, pVII, pVIII, pIX) are components of the phage particle.

Unlike the situation in phages with larger genomes, where temporal regulation of gene expression is the rule, filamentous phage proteins are synthesized concurrently. Diverse mechanisms ensure that each is produced at an appropriate rate. There are differences in promoter and ribosome binding site strength or accessibility (7, 8). Gene VII is translated from an inherently defective translation initiation site and is coupled to its upstream gene (76, 77, 182). A weak rho-dependent termination signal at the beginning of gene I limits its transcription, and a large number of infrequently used codons reduces the rate of its translation (70). At the other end of the spectrum, overlapping transcripts from multiple promoters (there are only two terminators) and multiple RNA processing events by RNaseE (157) increase the abundance of RNAs for the genes closest to the terminators (50). This results in high levels of pV and pVIII, the proteins that are required in the greatest quantities. IKe exhibits a similar pattern of overlapping mRNAs with a common 3′ end, but surprisingly, the only highly conserved regulatory element is the rho-independent terminator that generates one of the common 3′ ends (156).

The rates of phage protein and DNA synthesis slow at later times after infection, when high concentrations of pV sequester the (+) strands and prevent their conversion to RF. The reduced amount of RF (i.e., template) results in a lower level of gene expression. In addition, excess pV represses the translation of pII and pX, and the reduction in pII levels leads to even lower rates of (+) strand synthesis (44, 46). pV, which is a nonspecific single-stranded DNA binding protein, represses translation by binding specifically to a GT-rich tetraplex structure that is present in both geneII and geneX mRNAs, just upstream of their translation initiation sites (109, 115). The net result of these control systems (under optimal conditions) is that a steady-state level of phage DNA and proteins is maintained; synthesis is balanced by secretion of progeny phage and by the continued growth and division of the infected cells. Thus phage production continues indefinitely at a linear rate.

Phage Assembly

Filamentous phage assembly is a secretory process. Assembly occurs in the cytoplasmic membrane, and nascent phages are secreted from the cell as they assemble (figure 12-5). All eight of the phage-encoded proteins that are directly involved in assembly are integral membrane proteins. This includes three nonvirion proteins—pI and its restart partner, pXI, in the cytoplasmic membrane (69) and pIV in the outer membrane (12)—and the five viral coat proteins, which reside in the cytoplasmic membrane prior to their incorporation into phage (36). Two (pIII and pVIII) are synthesized as precursors and, after signal sequence cleavage, span the membrane with their C-termini in the cytoplasm. The orientation of the other coat proteins is not known with certainty. Like the major coat protein, pVII and pIX contain one or more N-terminal negatively charged residues and one or more C-terminal positively charged residues, which according to the "positive inside" rule articulated by von Heijne (161), suggests that, like pVIII, they span the membrane with their C-termini in the cytoplasm. Furthermore, successful display of proteins fused to the N-termini of pVII and pIX appears to require a signal sequence (48), which also suggests that the N-termini are periplasmic. pVI is particularly hydrophobic and could span the membrane more than once.

Progeny phage particles first appear in the culture supernatant about 10 minutes after infection at 37°C. Their numbers increase exponentially for about 40 minutes, after which the rate becomes linear. About 1000 phage per cell are produced in the first hour. Under optimal conditions, the infected cells can continue to grow and divide—and produce phage—indefinitely. But even modest perturbations of the phage life cycle can lead to the eventual death of infected cells and formation of clear rather than the normal turbid plaques (140). Most nonproductive infections result in death of the infected bacteria, presumably

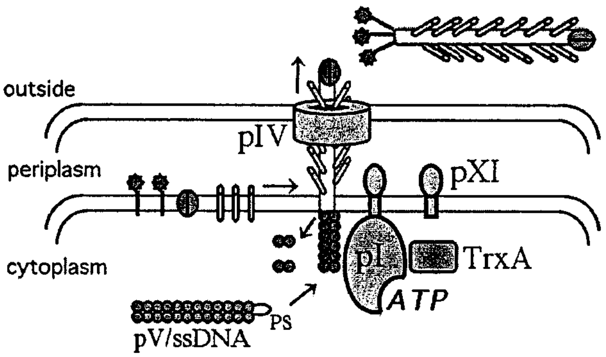


Figure 12-5 A hypothetical model for assembly and extrusion of filamentous phage. Initiation of phage assembly occurs when a cytoplasmic region of pI recognizes the packaging signal (PS), which protrudes from one end of the pV–ss DNA complex. Assembly occurs at sites where the cytoplasmic and outer membranes are brought together by trans-envelope interactions between pI and pXI in the cytoplasmic membrane and pIV in the outer membrane. Assembly starts with the addition of three to five copies each of two small "tip" proteins (pVII and pIX). During elongation, which requires host thioredoxin (TrxA) and ATP hydrolysis (presumably by pI), pV is stripped from the DNA and several thousand copies of the major coat protein, pVIII, are added to form the phage tube, which surrounds the single-stranded DNA. When the end of the DNA has been reached, pIII and pVI are added, and a conformational change in pIII detaches them from the membrane. The completed phage is thereby released, and extruded through the pIV channel.

due to the accumulation of phage DNA in the cytoplasm and/or capsid proteins in the cytoplasmic membrane. The exceptions are mutants that lack either pII or pIII. The absence of pII prevents phage DNA replication; as a consequence, little phage gene expression occurs and the phage DNA is lost by dilution. When pIII is absent, phage filaments are still efficiently assembled and extruded across the membrane but they remain attached to the cell; thus phage DNA does not accumulate in the cytoplasm and structural proteins do not accumulate in the cytoplasmic membrane (122).

Phage assembly occurs at distinct membrane assembly sites. Assembly sites have been visualized by electron microscopy as regions where the cytoplasmic and outer membranes are in close contact (97). The assembly site is a trans-envelope complex of the three phage-encoded morphogenetic proteins membrane; genetic (24, 136) and biochemical (40) results indicate that the C-terminal periplasmic domain of pI and pXI (cytoplasmic membrane-spanning proteins) interacts with the N-terminal periplasmic domain of pIV (an outer membrane protein). The assembly site forms independently of other phage proteins, and when its formation is blocked by mutation, pI becomes susceptible to cleavage by a host protease (40).

pIV forms a large multimer in the outer membrane that is composed of 12 to 14 identical subunits (83, 116, 134). Like the outer membrane porins, the multimer is extremely resistant to dissociation by detergents (94). As illustrated in cartoon form in figure 12-5 (but based on actual aligned images), the pIV multimer is somewhat barrel-shaped in side view; in top views, it is cylindrical with a central cavity (~8 nm in diameter) sufficient to accommodate an emerging phage (95). This is a much larger diameter than that of known outer membrane channel-forming proteins (86), but there appears to be some density within the cavity (116), which could explain why the presence of wild-type pIV does not disrupt the integrity of the outer membrane. Certain amino acid substitutions in pIV, however, do render bacteria hypersensitive to detergents and allow entry of foreign substances into the bacterial periplasm (102, 139). In planar lipid bilayers, one such mutant pIV forms highly conductive channels; wild-type pIV has similar conductivity, but its probability of being open is much lower than that of the mutant (102). Phage pass through the pIV channel, as shown by the ability of tethered phage (that lack pIII) to block the pIV-dependent diffusion of maltodextrins across the outer membrane (103). Nothing is known about the channel interior, except that it can accommodate heterologous pVIII from IKe (132), as well as an occasional pVIII subunit carrying a large N-terminal extension (81) or pVIII uniformly substituted with short (6–8 residue) N-terminal extensions (55, 74). The size and/or properties of the channel may be a limiting factor in phage display.

The N-terminal third of pIV forms a trypsin-resistant domain that extends into the periplasm (12). Although pIV from IKe cannot replace f1 pIV to assemble f1 phage (132), a chimeric IKe–f1 pIV could, showing that pIV specificity is determined by this domain (24). Genetic analysis (136) and the poor conservation of the periplasmic domains of pI/pXI from IKe and f1 suggest that the periplasmic domain of pIV interacts with pI/pXI. The C-terminal half of pIV mediates association with the membrane (138). Proteins homologous to this region of pIV (a span of about 200 residues) have been identified in many Gram-negative bacterial species and are called "secretins." Bacterial secretins are essential components in either of two different specialized protein secretion systems (135). Both type II and type III systems secrete proteins across both bacterial membranes and, like pIV, the homologs form multimeric rings (133, 154).

pI and pXI form an equimolar multimeric complex composed of about five or six copies of each (39); its shape and dimensions are not known. In the absence of the other phage proteins, pI/pXI causes membrane depolarization (69, 70), which suggests that the complex may also be a channel. pXI, the product of translational reinitiation at codon 241 in gene I, is required (in addition to pI) for phage production (124); however, despite the identity of the sequences they share, certain mutations that are tolerated in pXI do not support phage assembly if they are present in both pI and pXI (58). The cytoplasmic N-terminal domain of pI from all known filamentous phages contains a nucleotide-binding motif. The integrity of this motif is essential for f1 phage assembly (40). Furthermore, f1 phage assembly requires ATP hydrolysis (41). A likely inference is that pI is an ATPase. The pI proteins of Ff and IKe are not interchangeable (132), but if pI and pIV are swapped simultaneously, heterologous phage assembly occurs, indicating that only homologous protein pairs interact (136).

Particle assembly requires the assembly site (pI, pXI, and pIV), an appropriately presented single-stranded DNA molecule, the major coat protein, two minor coat proteins (pVII and pIX), and host-encoded thioredoxin. Genetic analyses suggest that pVII and pIX interact with the PS, which protrudes from the pV–ssDNA complex, that the PS also associates with the cytoplasmic domain of pI (140), and that pI interacts with thioredoxin, a small, cytoplasmic protein known as a potent reductant of protein disulfides (137). Thioredoxin may confer processivity to the elongation reaction, as it does in phage T7 DNA replication (6). Its redox activity is dispensable for both processes (141). Phage elongation involves the successive replacement of pV dimers by the membrane-embedded coat proteins and the simultaneous translocation of the DNA across the membrane. Perhaps the pI/pXI complex is a channel that the coat proteins must penetrate to gain access to the DNA; alternatively, the complex may be a "screw" that winds the DNA across the membrane, picking up coat proteins on the way.

Phage elongation continues until the end of the viral DNA has been coated by pVIII. If either pIII or pVI is absent, the largely extracellular phage particle remains tethered to the cytoplasmic membrane where it remains competent to resume elongation when another pV–ssDNA complex enters the assembly site; ultimately, tethered phage filaments 10 times or more the length of a normal phage particle accumulate (122). Even when pIII and pVI are present, about 5% of progeny phage particles are double length. Secondary rounds of elongation do not require reinitiation: single-stranded DNA without a PS can be efficiently incorporated as a passenger, behind DNA that does contain a PS (140).

Upon incorporation of the membrane-embedded pIII–pVI complex at the terminal end of the nascent phage particle, these proteins are thought to undergo a conformational change that stabilizes them and detaches them from the membrane (121). A fragment containing only the C-terminal 83 residues of pIII is sufficient to bind pVI and incorporate the incomplete dimer onto the particle, but it cannot detach the phage from the cell. Release of the phage requires a slightly longer (93 residue) C-terminal segment of pIII. A still longer portion of pIII (the 132 C-terminal residues) is required for the formation of stable virus particles.

Host Responses to Filamentous Phage Infection

Infection by filamentous phage induces a transient alteration in phospholipid metabolism; the rate of phosphatidylglycerol and cardiolipin synthesis increases and phosphatidylethanolamine synthesis decreases (178). This effect is due to the accumulation of pVIII in the membrane, which inhibits the activity of *E. coli* phosphatidylserine synthetase (17).

Infection also affects the phosphorylation level of several host proteins; phosphorylation of the heat shock protein DnaK and at least seven other proteins increases, while that of several other proteins declines (127).

Whether expressed as a consequence of phage infection or a cloned gene, pIV induces the high-level synthesis of a set of bacterial proteins called *p*hage *s*hock *p*roteins (13, 14). The *psp* genes are so named because their expression is induced by phage infection and by other shocks, such as heat and osmotic shock (14). Transcription of the *psp* genes is also induced by bacterial homologs of pIV (59, 174) and occurs as a consequence of improper localization (60). Similarly, when delivery of pIV to the outer membrane was improved, *psp* induction disappeared (23). How the signal is transduced is not known.

The *psp* genes (*psp*ABCE) are arranged in an operon controlled by $\sigma 54$ (111). Expression of the operon requires PspF, a constitutively active transcriptional activator encoded by a gene adjacent to and divergently transcribed from the *psp*ABCE operon (32, 80). By binding to the upstream activating sequences necessary to activate the *psp* operon, PspF obscures its own $\sigma 70$ promoter, thereby autogenously repressing its own synthesis (79). PspA negatively regulates the operon. Its failure to bind to DNA suggested that it might act by binding to PspF (31), and this has recently been confirmed (1). PspB and PspC cooperatively activate the operon, perhaps by antagonizing PspA-controlled repression (169). Both PspB and PspC bind to PspA in vitro, but PspB binds only when PspC is present (1). Transcription of *psp* is also dependent on IHF, which bends the DNA upstream of the promoter, thereby positioning PspF with respect to $\sigma 54$-RNA polymerase bound at the promoter (33, 170). The function of the Psp proteins, other than to regulate their own expression, is not clear. The *psp* operon is highly induced in late stationary phase (171) and when the Sec protein secretion system is blocked (84). Bacteria that lack the *psp*ABC genes show a decreased ability to survive in stationary phase under alkaline conditions (171). These observations suggest a role, probably for PspA, in maintenance of endogenous energy sources (84, 171). These mutant phenotypes are fairly subtle, but in *Yersinia enterocolitica*, the *psp* genes are required for virulence and for growth in vitro when the Ysc type III secretion system is produced (25).

Homologs of PspA have been identified in several higher plants, where they localize to the inner and thylakoid membranes of chloroplasts (89, 92); a disruption mutant of vipp1, which encodes the homolog, formed aberrantly structured thylakoid membranes (175).

Filamentous Phage Display

Phage display is an extraordinarily powerful method of linking phenotype and genotype in the selection of proteins or peptides (see also chapter 44). The principle is very simple: a DNA sequence is inserted into a gene encoding a phage coat protein so as to make a fusion protein in which the coat protein now "displays" the protein product of the inserted sequence. If a ligand is bound to a solid support, those phages expressing a protein or peptide that binds to the ligand will be immobilized, the remaining phages can be washed away, and the bound phage released and amplified by another cycle of growth. This process, called panning, can then be reiterated. Under the best of circumstances an enrichment of 10^4 to 10^5 can be achieved at every cycle, and usually four or five cycles of binding and regrowth suffice to produce one or a few selected phage carrying inserts of high affinity.

Most phage display utilizes libraries that contain at least some random sequences. Random peptides may be sandwiched in between a constant framework, constrained by paired cysteine residues, or placed in an exposed loop of a suitable protein scaffold. Often sequences of antibody combining sites are randomized. Antibody FABs can be expressed, either as single-chain antibodies (in which the two chains are covalently joined by a linker) or as individual chains (177). The other chain, if equipped with a signal sequence, can be transported into the periplasm, and there will pair with its partner to form an FAB (3).

This description is, of course, one for an idealized case. Although the method has worked as described in many instances, complications and difficulties show up, and many routes have been found to overcome or circumvent the problems. The utility of the method is such that it has become a cottage industry. Since its introduction by George Smith in 1985 (149), thousands of phage display papers have appeared in the literature, and both the methodology and the results are reviewed regularly.

Initial cloning was into gene III, which encodes the minor protein that helps to terminate assembly and that is required for infection of a new bacterial host. The original cloning sites were at or near the N-terminus (which is prominently displayed on the surface of the phage) at positions in which relatively small inserts do not interfere with any of the protein's functions (figure 12-6C). Larger proteins do interfere with function but the phages often are still infective. For pIII display of larger proteins, the chimeric protein is usually expressed from a plasmid that, besides its own origin and antibiotic resistance gene, bears a phage origin of replication and a packaging signal. When a cell carrying such a plasmid is infected with a helper phage, both phagemid and helper phage DNA are encapsidated and the progeny particles carry a mixture of the wild-type protein and the fusion protein (figure 12-6B). It is often desirable to control expression of the fusion protein so that at most one copy will be present in any particle; when this is the case the panning procedure will usually lead to the isolation of phages with the highest affinity to the target. When more than one copy of the fusion is expressed, avidity due to a range of multiplicities can lead to the isolation of binders of lower affinity.

Il'ichev et al. (75), to the surprise of most phage biologists, found that peptides could be added not only to the N-terminus of pIII but also to pVIII, the major coat protein, of which there are about 2700 copies in a wild-type Ff phage (figure 12-6B). Similar observations were subsequently reported by several other laboratories (38, 55, 104). When the fusion protein is the only source of phage coat, peptides of length 6 seem to be generally tolerated, while only a subset of longer peptides leave the coat functional (74). If wild-type coat protein is supplied, much longer peptides may be presented on the phage surface (55), and indeed even FABs are displayed, although at much lower copy number than peptides (81) (figure 12-6C).

Because of the very high copy number of pVIII peptide display, phage have been used as immunogens (55, 110, 181). They can also serve as tools with which to identify relatively low affinity peptides whose binding might not be strong enough to withstand the washing procedures. Sequences that encode weak binders are then often mutagenized and reselected, in the expectation, often achieved, that higher affinity binders will be found amongst the progeny (37, 180). Such methods attempt to recapitulate in principle the workings of the immune system, in which initial, relatively low affinity binders in immunoglobulins that form multimers are followed, after suitable selection, by binders of higher affinity carried on monomeric species.

There are constraints on display at the N-terminus of either pIII or pVIII. The fusions must be compatible with export through the host's inner membrane. The insertion in pIII or pVIII must be between a signal sequence and the mature part of the protein, and the inserts must be in frame with both the signal sequence and the mature protein and should not contain stop codons. When the sequence encoding the displayed protein is a synthetic construct, the frame and stop codon limitations may not be a problem, but if the inserted sequence is random, or comes from random DNA fragments, the phase and stop codon limitations are manifest, since for a random fragment only one eighteenth can be expected to be oriented correctly and in

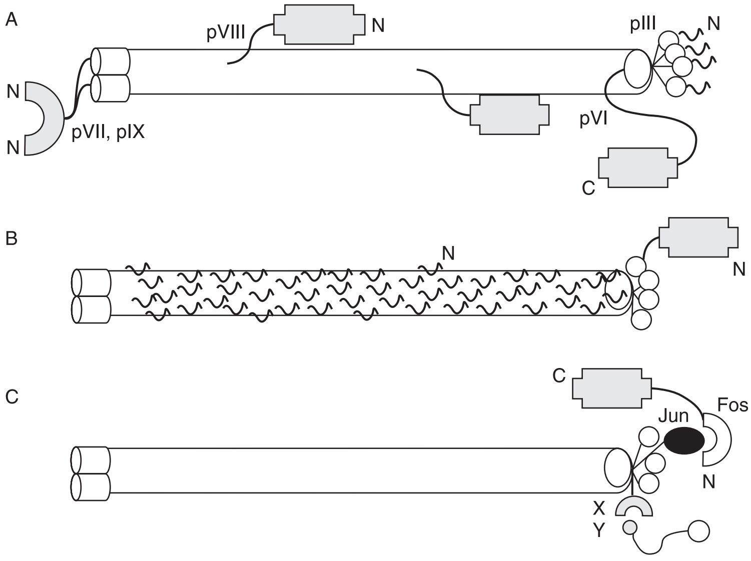


Figure 12-6 Outline of several filamentous phage display strategies. A: pVII and pIX display: interacting proteins fused to their N-termini (signal sequences upstream, wild-type pVII and pIX provided). pVIII display: large proteins fused to the N-terminus of pVIII displayed at low copy only (wild type pVIII provided). pVI display: proteins fused to the C-terminus displayed, though inefficiently (wild type pVI provided). pIII display: short peptides at the N-terminus of pIII can be uniformly displayed without disrupting infectivity. B: pVIII display: short peptides at the N-terminus of pVIII can be uniformly displayed without disrupting virus assembly; pIII display: large proteins fused to the N-terminus of pIII (wild-type pIII provided). C: Fos–Jun C-terminal display: a Fos–protein X_N fusion associates with N-terminally displayed Jun–pIII in the periplasm, and the entire sandwich incorporates into virions that contain the X_N DNA sequence. Selectively infective phage (SIP): protein X_a fused to the N-terminus of the C-domain of pIII and protein Y_a fused to the C-terminus of the N1–N2 domains of pIII; X_aY_a interaction makes particles infectious. See chapter 44 for additional discussion of phage-mediated protein display.

frame with both the signal sequence and the downstream phage gene.

A number of ways, some very ingenious, have been found that avoid these problems. In order to make fusions with a free C-terminus, Crameri and Suter (22) used Fos and Jun, which bind tightly, as an adapter pair (figure 12-6C). They constructed a vector that expresses a Jun–pIII fusion and carries a cloning site arranged so that polypeptides encoded by cloned DNA are fused to the C-terminus of Fos. Both Fos and Jun were engineered to contain unpaired cysteine residues, which form disulfides and prevent exchange between different Jun–Fos pairs. This permitted display of proteins from cDNA libraries, which often contain stop codons that, if inserted into gene III, would prevent expression of its C-terminal region.

Jespers et al. (78) found that fusions to the C-terminus of pVI were accessible, albeit that the presentation was much less efficient (circa 10^{-3}) than fusion of the same epitope to the N-terminus of pIII (figure 12-6C). More recently, Sidhu and coworkers have successfully cloned on the C-terminus of pIII (43). There is some evidence that in wild-type phage the C-terminus of pVI is not available to antiserum (36); it is also clear from proteolytic and other studies that at least most of the C-terminus of pIII is embedded in the phage particle (2, 113, 122). Thus these surprising results probably reflect much more structural plasticity in the phage than had previously been suspected, rather than indicating where the actual C-termini of these proteins are in the normal virion.

The plasticity hypothesis (or perhaps degeneracy) is supported by work that showed that the requirements for the incorporation of a pVIII fusion protein are quite limited and that very extensive mutagenesis is compatible with display and in some instances can much improve it (146, 173). These observations led to the construction of a "minimal" coat protein that retains only nine non-Ala residues and yet is incorporated into phage (129). A "reversed" major coat protein was also constructed, in which the pattern of residues found to be necessary from the mutagenesis studies was inverted in the N-to-C-terminal direction (172). Both the "minimal" and "reversed" proteins were selected to display a ligand. Neither the "minimal" nor the "reversed" protein is, by itself, sufficient to support phage assembly; they require a source of wild-type (or close to wild-type) coat protein. Hence, as had been previously noted in a different context (55), the requirements for incorporation of some coat monomers into phage are simpler than those governing complete assembly.

Gao et al. (48) have demonstrated that by adding a signal sequence and linkers to pVII and pIX, the two very small proteins at one end of the phage, proteins can be fused to their N-termini and displayed on the phage surface (figure 12-6A). When a V_L sequence was fused to gene IX, and a V_H to gene VII, the phages carrying these fusion proteins bound to immobilized antigen and also had catalytic antibody activity (48). In control experiments, phage carrying only one of the fusions did not bind the antigen. These experiments establish the possibility of displaying heterodimeric arrays such as FABs, confirm the location and orientation of pIX, and establish the localization and orientation of pVII in the virion.

Display has also been adapted to the identification of pairs of interacting proteins or peptides, even though they may both be unknown. The underlying principle is that the N- and C-terminal parts of pIII are expressed from different vectors or different sites in one vector. Each half of pIII is fused to an interaction candidate. Only if the two parts are brought together by interaction between the interaction candidates will the resulting phage be infective (29) (reviewed by Spada et al. 151).

Phage display has been used in varied ways, too numerous to cite in any detail. The two most frequent uses have been to find peptides that bind to particular ligands, and to search for or generate antibodies with desirable characteristics. Phages decorated with peptides have been used as specific immunogens. Antibodies of very high affinity have been generated from either immunized or naïve libraries. Linking the phage to a support and requiring cleavage of a peptide for its release has defined protease substrates. Hormones of higher or altered affinity have been selected. Catalytic antibodies have been developed. Recently, displayed peptides have been used for "in vivo" phage display, in which a library of peptide fusions is injected into a mammal, and phage which are found in particular organs or blood vessels are isolated, in the expectation that they provide peptides that can "home in" on their target (118, 131).

There are more than 100 reviews that summarize display work, including chapter 44 in this book. We cite some relatively recent ones (16, 21, 42, 57, 65, 68, 85, 105, 117, 120, 144, 166, 177), but there are many others. The book by Barbas et al. (4) provides help with the methodology, as do a number of papers in *Methods in Enzymology* (91, 114, 125, 128, 145, 150, 160, 165, 176, 179).

References

1. Adams, H. 2002. Protein export in *Escherichia coli*: involvement of the signal sequence and Psp proteins. PhD thesis. Utrecht University, Utrecht, The Netherlands.
2. Armstrong, J., R. N. Perham, and J. E. Walker. 1981. Domain-structure of bacteriophage fd adsorption protein. FEBS Lett. 135:167–172.
3. Barbas, C. F. 3rd, A. S. Kang, R. A. Lerner, and S. J. Benkovic. 1991. Assembly of combinatorial antibody libraries on phage surfaces: the gene III site. Proc. Nat. Acad. Sci. USA 88:7978–7982.
4. Barbas C. F. III, D. R. Burton, J. K. Scott, and G. J. Silverman. 2001. Phage Display: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
5. Bauer, M., and G. P. Smith. 1988. Filamentous phage morphogenetic signal sequence and orientation of DNA in the virion and gene V protein complex. Virology 167:166–175.
6. Bedford, E., S. Tabor, and C. C. Richardson. 1997. The thioredoxin binding domain of bacteriophage T7 DNA polymerase confers processivity on *Escherichia coli* DNA polymerase I. Proc. Nat. Acad. Sci. USA 94:479–484.
7. Blumer, K. J., M. R. Ivey, and D. A. Steege. 1987. Translational control of phage f1 gene expression by differential activities of the gene V, VII, IX and VIII initiation sites. J. Mol. Biol. 197:439–451.
8. Blumer, K. J., and D. A. Steege. 1989. Recognition and cleavage signals for messenger-RNA processing lie within local domains of the phage f1 RNA precursors. J. Biol. Chem. 264:20770–20777.
9. Boeke, J. D., P. Model, and N. D. Zinder. 1982. Effects of bacteriophage f1 gene III protein on the host cell membrane. Mol. Gen. Genet. 186:185–192.
10. Boyd, E. F., K. E. Moyer, L. Shi, and M. K. Waldor. 2000. Infectious CTXPhi and the *Vibrio* pathogenicity island prophage in *Vibrio mimicus*: evidence for recent horizontal transfer between *V. mimicus* and *V. cholerae*. Infect. Immun. 68:1507–1513.
11. Bradley, D. E., and J. Whelan. 1989. *Escherichia coli tolQ* mutants are resistant to filamentous bacteriophages that adsorb to the tips, not the shafts, of conjugative pili. J. Gen. Microbiol. 135:1857–1863.
12. Brissette, J. L., and M. Russel. 1990. Secretion and membrane integration of a filamentous phage-encoded morphogenetic protein. J. Mol. Biol. 211:565–580.
13. Brissette, J. L., M. Russel, L. Weiner, and P. Model. 1990. Phage shock protein, a stress protein of *Escherichia coli*. Proc. Natl. Acad. Sci. USA 87:862–866.
14. Brissette, J. L., L. Weiner, T. L. Ripmaster, and P. Model. 1991. Characterization and sequence of the *Escherichia coli* stress-induced *psp* operon. J. Mol. Biol. 220:35–48.
15. Bross, P., K. Bussmann, W. Keppner, and I. Rasched. 1988. Functional analysis of the adsorption protein of two filamentous phages with different host specificities. J. Gen. Microbiol. 134:461–471.
16. Cesareni, G., L. Castagnoli, and G. Cestra. 1999. Phage displayed peptide libraries. Combinatorial Chemistry & High Throughput Screening 2:1–17.
17. Chamberlain, B. K., and R. E. Webster. 1978. Effect of membrane-associated f1 bacteriophage coat protein upon the activity of *Escherichia coli* phosphatidylserine synthetase. J. Bacteriol. 135:883–887.
18. Chopin, M. C., A. Rouault, S. D. Ehrlich, and M. Gautier. 2002. Filamentous phage active on the gram-positive bacterium *Propionibacterium freudenreichii*. J. Bacteriol. 184:2030–2033.

19. Click, E. M., and R. E. Webster. 1997. Filamentous phage infection: required interactions with the TolA protein. J. Bacteriol. 179:6464–6471.
20. Click, E. M., and R. E. Webster. 1998. The TolQRA proteins are required for membrane insertion of the major capsid protein of the filamentous phage f1 during infection. J. Bacteriol. 180:1723–1728.
21. Cortese, R., P. Monaci, A. Luzzago, C. Santini, F. Bartoli, I. Cortese, P. Fortugno, G. Galfre, A. Nicosia, and F. Felici. 1996. Selection of biologically active peptides by phage display of random peptide libraries. Curr. Opin. Biotechnol. 7:616–621.
22. Crameri, R., and M. Suter. 1993. Display of biologically active proteins on the surface of filamentous phages: a cDNA cloning system for selection of functional gene products linked to the genetic information responsible for their production. Gene 137:69–75.
23. Daefler, S., I. Guilvout, K. R. Hardie, A. P. Pugsley, and M. Russel. 1997. The C-terminal domain of the secretin PulD contains the binding site for its cognate chaperone, PulS, and confers PulS dependence on pIVf1 function. Mol. Microbiol. 24:465–475.
24. Daefler, S., M. Russel, and P. Model. 1997. Module swaps between related translocator proteins pIV(f1), pIV(IKe) and PulD: identification of a specificity domain. J. Mol. Biol. 266:978–992.
25. Darwin, A. J., and V. L. Miller. 2001. The psp locus of *Yersinia enterocolitica* is required for virulence and for growth in vitro when the Ysc type III secretion system is produced. Mol. Microbiol. 39:429–444.
26. Davis, B. M., E. H. Lawson, M. Sandkvist, A. Ali, S. Sozhamannan, and M. K. Waldor. 2000. Convergence of the secretory pathways for cholera toxin and the filamentous phage, CTXphi. Science 288:333–335.
27. Deng, L.-W., and R. N. Perham. 2002. Delineating the site of interaction on the pIII protein of filamentous bacteriophage fd with the F-pilus of *Escherichia coli*. J. Mol. Biol. 319:603–614.
28. Deng, L. W., P. Malik, and R. N. Perham. 1999. Interaction of the globular domains of pIII protein of filamentous bacteriophage fd with the F-pilus of *Escherichia coli*. Virology 253:271–277.
29. Duenas, M., and C. A. K. Borrebaeck. 1994. Clonal selection and amplification of phage display antibodies by lining antigen recognition and phage replication. Bio/Technology 12:999–1002.
30. Dunker, A. K., L. D. Ensign, G. E. Arnold, and L. M. Roberts. 1991. A model for fd phage penetration and assembly. FEBS Lett. 292:271–274.
31. Dworkin, J., G. Jovanovic, and P. Model. 2000. The PspA protein of *Escherichia coli* is a negative regulator of sigma(54)-dependent transcription. J. Bacteriol. 182:311–319.
32. Dworkin, J., G. Jovanovic, and P. Model. 1997. Role of upstream activation sequences and integration host factor in transcriptional activation by the constitutively active prokaryotic enhancer-binding protein PspF. J. Mol. Biol. 273:377–388.
33. Dworkin, J., A. J. Ninfa, and P. Model. 1998. A protein-induced DNA bend increases the specificity of a prokaryotic enhancer-binding protein. Genes Dev. 12:894–900.
34. Endemann, H., P. Bross, and I. Rasched. 1992. The adsorption protein of phage IKe. Localization by deletion mutagenesis. Mol. Microbiol. 6:471–478.
35. Endemann, H., V. Gailus, and I. Rasched. 1993. Interchangeability of the adsorption proteins of bacteriophages Ff and IKe. J. Virol. 67:3332–3337.
36. Endemann, H., and P. Model. 1995. Location of filamentous phage minor coat proteins in phage and in infected cells. J. Mol. Biol. 250:496–506.
37. Fairbrother, W. J., H. W. Christinger, A. G. Cochran, G. Fuh, C. J. Keenan, C. Quan, S. K. Shriver, J. Y. Tom, J. A. Wells, and B. C. Cunningham. 1998. Novel peptides selected to bind vascular endothelial growth factor target the receptor-binding site. Biochemistry 37:17754–17764.
38. Felici, F., L. Castagnoli, A. Musacchio, R. Jappelli, and G. Cesareni. 1991. Selection of antibody ligands from a large library of oligopeptides expressed on a multivalent exposition vector. J. Mol. Biol. 222:301–310.
39. Feng, J. N., P. Model, and M. Russel. Unpublished data.
40. Feng, J. N., P. Model, and M. Russel. 1999. A trans-envelope protein complex needed for filamentous phage assembly and export. Mol. Microbiol. 34:745–755.
41. Feng, J. N., M. Russel, and P. Model. 1997. A permeabilized cell system that assembles filamentous bacteriophage. Proc. Natl. Acad. Sci. USA 94:4068–4073.
42. Forrer, P., S. Jung, and A. Pluckthun. 1999. Beyond binding: using phage display to select for structure, folding and enzymatic activity in proteins. Curr. Opin. Struct. Biol. 9:514–520.
43. Fuh, G., and S. S. Sidhu. 2000. Efficient phage display of polypeptides fused to the carboxy-terminus of the M13 gene-3 minor coat protein. FEBS Lett. 480:231–234.
44. Fulford, W., and P. Model. 1988. Bacteriophage f1 DNA replication genes. II. The roles of gene V protein and gene II protein in complementary strand synthesis. J. Mol. Biol. 203:39–48.
45. Fulford, W., and P. Model. 1984. Gene X of bacteriophage f1 is required for phage DNA synthesis. Mutagenesis of in-frame overlapping genes. J. Mol. Biol. 178:137–153.
46. Fulford, W., and P. Model. 1988. Regulation of bacteriophage f1 DNA replication. I. New functions for genes II and X. J. Mol. Biol. 203:49–62.
47. Gailus, V., and I. Rasched. 1994. The adsorption protein of bacteriophage fd and its neighbour minor coat protein build a structural entity. Eur. J. Biochem. 222:927–931.
48. Gao, C., S. Mao, C. H. Lo, P. Wirsching, R. A. Lerner, and K. D. Janda. 1999. Making artificial antibodies: a format for phage display of combinatorial heterodimeric arrays. Proc. Natl. Acad. Sci. USA 96:6025–6030.
49. Gaspar, J. A., J. A. Thomas, C. L. Marolda, and M. A. Valvano. 2000. Surface expression of O-specific lipopolysaccharide in *Escherichia coli* requires the function of the TolA protein. Mol. Microbiol. 38:262–275.
50. Goodrich, A. F., and D. A. Steege. 1999. Roles of polyadenylation and nucleolytic cleavage in the filamentous phage

mRNA processing and decay pathways in *Escherichia coli*. RNA 5:972–985.

51. Gray, C.W. 1989. Three-dimensional structure of complexes of single-stranded DNA-binding proteins with DNA. IKe and fd gene 5 proteins form left-handed helices with single-stranded DNA. J. Mol. Biol. 208:57–64.
52. Gray, C.W., R. S. Brown, and D. A. Marvin. 1981. Adsorption complex of filamentous fd virus. J. Mol. Biol. 146:621–627.
53. Greenstein, D., and K. Horiuchi. 1989. Double-strand cleavage and strand joining by the replication initiator protein of filamentous phage f1. J. Biol. Chem. 264:12627–12632.
54. Greenwood, J., G. J. Hunter, and R. N. Perham. 1991. Regulation of filamentous bacteriophage length by modification of electrostatic interactions between coat protein and DNA. J. Mol. Biol. 217:223–227.
55. Greenwood, J., A. E. Willis, and R. N. Perham. 1991. Multiple display of foreign peptides on a filamentous bacteriophage. J. Mol. Biol. 220:821–827.
56. Griffith, J., M. Manning, and K. Dunn. 1981. Filamentous bacteriophage contract into hollow spherical-particles upon exposure to a chloroform–water interface. Cell 23:747–753.
57. Griffiths, A. D., and A. R. Duncan. 1998. Strategies for selection of antibodies by phage display. Curr. Opin. Biotechnol. 9:102–108.
58. Haigh, N. G., and R. E. Webster. 1999. The pI and pXI assembly proteins serve separate and essential roles in filamentous phage assembly. J. Mol. Biol. 293:1017–1027.
59. Hardie, K. R., S. Lory, and A. P. Pugsley. 1996. Insertion of an outer membrane protein in *Escherichia coli* requires a chaperone-like protein. EMBO J. 15:978–988.
60. Hardie, K. R., A. Seydel, I. Guilvout, and A. P. Pugsley. 1996. The secretin-specific, chaperone-like protein of the general secretory pathway: separation of proteolytic protection and piloting functions. Mol. Microbiol. 22:967–976.
61. Higashitani, A., D. Greenstein, H. Hirokawa, S. Asano, and K. Horiuchi. 1994. Multiple DNA conformational changes induced by an initiator protein precede the nicking reaction in a rolling circle replication origin. J. Mol. Biol. 237:388–400.
62. Higashitani, A., N. Higashitani, and K. Horiuchi. 1997. Minus-strand origin of filamentous phage versus transcriptional promoters in recognition of RNA polymerase. Proc. Natl. Acad. Sci. USA 94:2909–2914.
63. Higashitani, N., A. Higashitani, Z. W. Guan, and K. Horiuchi. 1996. Recognition mechanisms of the minus-strand origin of phage f1 by *Escherichia coli* RNA polymerase. Genes Cells 1:829–841.
64. Higashitani, N., A. Higashitani, and K. Horiuchi. 1993. Nucleotide sequence of the primer RNA for DNA replication of filamentous bacteriophages. J. Virol. 67:2175–2181.
65. Hoess, R. H. 2001. Protein design and phage display. Chem. Rev. 101:3205–3218.
66. Holliger, P., and L. Riechmann. 1997. A conserved infection pathway for filamentous bacteriophages is suggested by the structure of the membrane penetration domain of the minor coat protein g3p from phage fd. Structure 5:265–275.
67. Holliger, P., L. Riechmann, and R. L. Williams. 1999. Crystal structure of the two N-terminal domains of g3p from filamentous phage fd at 1.9 A: evidence for conformational lability. J. Mol. Biol. 288:649–657.
68. Hoogenboom, H. R., and P. Chames. 2000. Natural and designer binding sites made by phage display technology. Immunology Today 21:371–378.
69. Horabin, J. I., and R. E. Webster. 1988. An amino acid sequence which directs membrane insertion causes loss of membrane potential. J. Biol. Chem. 263:11575–11583.
70. Horabin, J. I., and R. E. Webster. 1986. Morphogenesis of f1 filamentous bacteriophage. Increased expression of gene I inhibits bacterial growth. J. Mol. Biol. 188:403–413.
71. Horiuchi, K. 1983. Co-evolution of a filamentous bacteriophage and its defective interfering particles. J. Mol. Biol. 169:389–407.
72. Horiuchi, K. 1997. Initiation mechanisms in replication of filamentous phage DNA. Genes Cells 2:425–432.
73. Hunter, G. J., D. H. Rowitch, and R. N. Perham. 1987. Interactions between DNA and coat protein in the structure and assembly of filamentous bacteriophage fd. Nature 327:252–254.
74. Iannolo, G., O. Minenkova, R. Petruzzelli, and G. Cesareni. 1995. Modifying filamentous phage capsid: limits in the size of the major capsid protein. J. Mol. Biol. 248:835–844.
75. Il'ichev, A. A., O. O. Minenkova, S. I. Tat'kov, N. N. Karpyshev, A. M. Eroshkin, V. I. Ofitserov, Z. A. Akimenko, V. A. Petrenko, and L. S. Sandakhchiev. 1990. The use of filamentous phage M13 in protein engineering [Rus]. Molekuliarnaia Biologiia 24:530–535.
76. Ivey-Hoyle, M., and D. A. Steege. 1992. Mutational analysis of an inherently defective translation initiation site. J. Mol. Biol. 224:1039–1054.
77. Ivey-Hoyle, M., and D. A. Steege. 1989. Translation of phage f1 gene VII occurs from an inherently defective initiation site made functional by coupling. J. Mol. Biol. 208:233–244.
78. Jespers, L. S., J. H. Messens, A. De Keyser, D. Eeckhout, I. Van Den Brande, Y. G. Gansemans, M. J. Lauwereys, G. P. Vlasuk, and P. E. Stanssens. 1995. Surface expression and ligand-based selection of cDNAs fused to filamentous phage gene VI. BioTechnology 13:378–382.
79. Jovanovic, G., J. Dworkin, and P. Model. 1997. Autogenous control of PspF, a constitutively active enhancer-binding protein of *Escherichia coli*. J. Bacteriol. 179:5232–5237.
80. Jovanovic, G., L. Weiner, and P. Model. 1996. Identification, nucleotide sequence, and characterization of PspF, the transcriptional activator of the *Escherichia coli* stress-induced *psp* operon. J. Bacteriol. 178:1936–1945.
81. Kang, A. S., C. F. Barbas, K. D. Janda, S. J. Benkovic, and R. A. Lerner. 1991. Linkage of recognition and replication functions by assembling combinatorial antibody Fab libraries along phage surfaces. Proc. Natl. Acad. Sci. USA 88:4363–4366.
82. Kay, B., G. Winter, and J. McCafferty (eds.). 1996. Biology of the Filamentous Bacteriophage. Academic Press, New York.
83. Kazmierczak, B. I., D. L. Mielke, M. Russel, and P. Model. 1994. pIV, a filamentous phage protein that mediates phage export across the bacterial cell envelope, forms a multimer. J. Mol. Biol. 238:187–198.

84. Kleerebezem, M., and J. Tommassen. 1993. Expression of the *pspA* gene stimulates efficient protein export in *Escherichia coli*. Mol. Microbiol. 7:947–956.

85. Klug, A. 1999. Zinc finger peptides for the regulation of gene expression. J. Mol. Biol. 293:215–218.

86. Koebnik, R., K. P. Locher, and P. Van Gelder. 2000. Structure and function of bacterial outer membrane proteins: barrels in a nutshell. Mol. Microbiol. 37:239–253.

87. Kokoska, R. J., and D. A. Steege. 1998. Appropriate expression of filamentous phage f1 DNA replication genes II and X requires RNase E-dependent processing and separate mRNAs. J. Bacteriol. 180:3245–3249.

88. Konings, R. N. H., R. H. A. Folmer, P. J. M. Folkers, M. Nilges, and C. W. Hilbers. 1995. Three-dimensional structure of the single-stranded DNA-binding protein encoded by gene V of the filamentous bacteriophage m13 and a model of its complex with single-stranded DNA. FEMS Microbiol. Rev. 17:57–72.

89. Kroll, D., K. Meierhoff, N. Bechtold, M. Kinoshita, S. Westphal, U. C. Vothknecht, J. Soll, and P. Westhoff. 2001. VIPP1, a nuclear gene of *Arabidopsis thaliana* essential for thylakoid membrane formation. Proc. Natl. Acad. Sci. USA 98:4238–4242.

90. Kuhn, A. 1995. Major coat proteins of bacteriophage Pf3 and M13 as model systems for Sec-independent protein transport. FEMS Microbiol. Rev. 17:185–190.

91. Laird-Offringa, I. A., and J. G. Belasco. 1996. In vitro genetic analysis of RNA-binding proteins using phage display libraries. Methods Enzymol. 267:149–168.

92. Li, H. M., Y. Kaneko, and K. Keegstra. 1994. Molecular cloning of a chloroplastic protein associated with both the envelope and thylakoid membranes. Plant Mol. Biol. 25:619–632.

93. Lin, N. T., B. Y. You, C. Y. Huang, C. W. Kuo, F. S. Wen, J. S. Yang, and Y. H. Tseng. 1994. Characterization of two novel filamentous phages of *Xanthomonas*. J. Gene. Virol. 75:2543–2547.

94. Linderoth, N. A., P. Model, and M. Russel. 1996. Essential role of a sodium dodecyl sulfate-resistant protein IV multimer in assembly-export of filamentous phage. J. Bacteriol. 178:1962–1970.

95. Linderoth, N. A., M. N. Simon, and M. Russel. 1997. The filamentous phage pIV multimer visualized by scanning transmission electron microscopy. Science 278:1635–1638.

96. Lloubes, R., E. Cascales, A. Walburger, E. Bouveret, C. Lazdunski, A. Bernadac, and L. Journet. 2001. The Tol-Pal proteins of the *Escherichia coli* cell envelope: an energized system required for outer membrane integrity? Res. Microbiol. 152:523–529.

97. Lopez, J., and R. E. Webster. 1985. Assembly site of bacteriophage f1 corresponds to adhesion zones between the inner and outer membranes of the host cell. J. Bacteriol. 163:1270–1274.

98. Lopez, J., and R. E. Webster. 1983. Morphogenesis of filamentous bacteriophage f1: orientation of extrusion and production of polyphage. Virology 127:177–193.

99. Lubkowski, J., F. Hennecke, A. Pluckthun, and A. Wlodawer. 1999. Filamentous phage infection: crystal structure of g3p in complex with its coreceptor, the C-terminal domain of TolA. Structure with Folding & Design 7:711–722.

100. Lubkowski, J., F. Hennecke, A. Pluckthun, and A. Wlodawer. 1998. The structural basis of phage display elucidated by the crystal structure of the N-terminal domains of g3p. Nature Struct. Biol. 5:140–147.

101. Makowski, L., and M. Russel. 1997. Structure and assembly of filamentous bacteriophages, pp. 352–380. *In* W. Chiu, R. M. Burnett, and R. L. Garcea (eds.) Structural Biology of Viruses. Oxford University Press, New York.

102. Marciano, D. K., M. Russel, and S. M. Simon. 1999. An aqueous channel for filamentous phage export. Science 284:1516–1519.

103. Marciano, D. K., M. Russel, and S. M. Simon. 2001. Assembling filamentous phage occlude pIV channels. Proc. Natl. Acad. Sci. USA 98:9359–9364.

104. Markland, W., B. L. Roberts, M. J. Saxena, S. K. Guterman, and R. C. Ladner. 1991. Design, construction and function of a multicopy display vector using fusions to the major coat protein of bacteriophage M13. Gene 109:13–19.

105. Marks, J. D., H. R. Hoogenboom, A. D. Griffiths, and G. Winter. 1992. Molecular evolution of proteins on filamentous phage. Mimicking the strategy of the immune system. J. Biol. Chem. 267:16007–16010.

106. Marvin, D. A. 1998. Filamentous phage structure, infection and assembly. Curr. Opin. Struct. Biol. 8:150–158.

107. Marvin, D. A., R. D. Hale, C. Nave, and M. H. Citterich. 1994. Molecular models and structural comparisons of native and mutant class I filamentous bacteriophages: Ff (fd, f1, M13), If1 and IKe. J. Mol. Biol. 235:260–286.

108. Michel, B., and N. D. Zinder. 1989. In vitro binding of the bacteriophage f1 geneV protein to the geneII RNA operator and its DNA analog. Nucleic Acids Res. 17:7333–7344.

109. Michel, B., and N. D. Zinder. 1989. Translational repression in bacteriophage f1: characterization of the gene V protein target on the gene II mRNA. Proc. Natl. Acad. Sci. USA 86:4002–4006.

110. Minenkova, O. O., A. A. Ilyichev, G. P. Kishchenko, and V. A. Petrenko. 1993. Design of specific immunogens using filamentous phage as the carrier. Gene 128:85–88.

111. Model, P., G. Jovanovic, and J. Dworkin. 1997. The *Escherichia coli* phage-shock-protein (*psp*) operon. Mol. Microbiol. 24:255–261.

112. Model, P., and M. Russel. 1988. Filamentous bacteriophage, pp. 375–456. *In* R. Calendar (ed.) The Bacteriophages. Plenum Publishing, New York.

113. Nelson, F. K., S. M. Friedman, and G. P. Smith. 1981. Filamentous phage DNA cloning vectors: a non-infective mutant with a non-polar deletion in gene III. Virology 108:338–350.

114. O'Neil, K. T., W. F. DeGrado, S. A. Mousa, N. Ramachandran, and R. H. Hoess. 1994. Identification of recognition sequences of adhesion molecules using phage display technology. Methods Enzymol. 245:370–386.

115. Oliver, A. W., I. Bogdarina, E. Schroeder, I. A. Taylor, and G. G. Kneale. 2000. Preferential binding of fd gene 5

protein to tetraplex nucleic acid structures. J. Mol. Biol. 301:575–584.

116. Opalka, N., R. Beckmann, N. Boisset, M. Simon, M. Russel, and S. Darst. 2003. Structure of the filamentous phage pIV multimer by cryo-electron microscopy. J. Mol. Biol. 325:461–470.

117. Parmley, S. F., and G. P. Smith. 1989. Filamentous fusion phage cloning vectors for the study of epitopes and design of vaccines. Adv. Exp. Med. Biol. 251:215–218.

118. Pasqualini, R., and E. Ruoslahti. 1996. Organ targeting in vivo using phage display peptide libraries. Nature 380:364–366.

119. Pederson, D. M., L. C. Welsh, D. A. Marvin, M. Sampson, R. N. Perham, M. Yu, and M. R. Slater. 2001. The protein capsid of filamentous bacteriophage PH75 from *Thermus thermophilus*. J. Mol. Biol. 309:401–421.

120. Perham, R. N., T. D. Terry, A. E. Willis, J. Greenwood, F. di Marzo Veronese, and E. Appella. 1995. Engineering a peptide epitope display system on filamentous bacteriophage. FEMS Microbiol. Rev. 17:25–31.

121. Rakonjac, J., J. N. Feng, and P. Model. 1999. Filamentous phage are released from the bacterial membrane by a two-step mechanism involving a short C-terminal fragment of pIII. J. Mol. Biol. 289:1253–1265.

122. Rakonjac, J., and P. Model. 1998. Roles of pIII in filamentous phage assembly. J. Mol. Biol. 282:25–41.

123. Rampf, B., P. Bross, T. Vocke, and I. Rasched. 1991. Release of periplasmic proteins induced in *E. coli* by expression of an N-terminal proximal segment of the phage fd gene 3 protein. FEBS Lett. 280:27–31.

124. Rapoza, M. P., and R. E. Webster. 1995. The products of gene I and the overlapping in-frame gene XI are required for filamentous phage assembly. J. Mol. Biol. 248:627–638.

125. Rebar, E. J., H. A. Greisman, and C. O. Pabo. 1996. Phage display methods for selecting zinc finger proteins with novel DNA-binding specificities. Methods Enzymol. 267:129–149.

126. Riechmann, L., and P. Holliger. 1997. The C-terminal domain of TolA is the coreceptor for filamentous phage infection of *E. coli*. Cell 90:351–360.

127. Rieul, C., J. C. Cortay, F. Bleicher, and A. J. Cozzone. 1987. Effect of bacteriophage M13 infection on phosphorylation of dnaK protein and other *Escherichia coli* proteins. Eur. J. Biochem. 168:621–627.

128. Roberts, B. L., W. Markland, and R. C. Ladner. 1996. Affinity maturation of proteins displayed on surface of M13 bacteriophage as major coat protein fusions. Methods Enzymol. 267:68–82.

129. Roth, T., G. Weiss, C. Eigenbrot, and S. S. Sidhu. 2002. A minimized m13 coat protein defines the minimum requirements for assembly into the bacteriophage particle. J. Mol. Biol. 322: 357–367.

130. Rowitch, D. H., G. J. Hunter, and R. N. Perham. 1988. Variable electrostatic interaction between DNA and coat protein in filamentous bacteriophage assembly. J. Mol. Biol. 204:663–674.

131. Ruoslahti, E. 2000. Targeting tumor vasculature with homing peptides from phage display. Semin. Cancer Biol. 10:435–442.

132. Russel, M. 1992. Interchangeability of related proteins and autonomy of function. The morphogenetic proteins of filamentous phage f1 and IKe cannot replace one another. J. Mol. Biol. 227:453–462.

133. Russel, M. 1998. Macromolecular assembly and secretion across the bacterial cell envelope: type II protein secretion systems. J. Mol. Biol. 279:485–499.

134. Russel, M. 1994. Mutants at conserved positions in gene IV, a gene required for assembly and secretion of filamentous phage. Mol. Microbiol. 14:357–369.

135. Russel, M. 1994. Phage assembly: a paradigm for bacterial virulence factor export? Science 265:612–614.

136. Russel, M. 1993. Protein–protein interactions during filamentous phage assembly. J. Mol. Biol. 231:689–697.

137. Russel, M. 1995. Thioredoxin genetics, pp. 265–275. *In* L. Packer (ed.) Biothiols, Part B: Glutathione and Thioredoxin Thiols in Signal Transduction and Gene Regulation. Methods in Enzymology, vol. 252. Academic Press, New York.

138. Russel, M., and B. Kazmierczak. 1993. Analysis of the structure and subcellular location of filamentous phage pIV. J. Bacteriol. 175:3998–4007.

139. Russel, M., N. A. Linderoth, and A. Sali. 1997. Filamentous phage assembly: variation on a protein export theme. Gene 192:23–32.

140. Russel, M., and P. Model. 1989. Genetic analysis of the filamentous bacteriophage packaging signal and of the proteins that interact with it. J. Virol. 63:3284–3295.

141. Russel, M., and P. Model. 1986. The role of thioredoxin in filamentous phage assembly. Construction, isolation, and characterization of mutant thioredoxins. J. Biol. Chem. 261:14997–15005.

142. Russel, M., H. Whirlow, T. P. Sun, and R. E. Webster. 1988. Low-frequency infection of F- bacteria by transducing particles of filamentous bacteriophages. J. Bacteriol. 170:5312–5316.

143. Sandkvist, M., M. Bagdasarian, and S. P. Howard. 2000. Characterization of the multimeric Eps complex required for cholera toxin secretion. Int. J. Med. Microbiol. 290:345–350.

144. Sidhu, S. S. 2001. Engineering M13 for phage display. Biomol. Eng. 18:57–63.

145. Sidhu, S. S., H. B. Lowman, B. C. Cunningham, and J. A. Wells. 2000. Phage display for selection of novel binding peptides. Methods Enzymol. 328:333–363.

146. Sidhu, S. S., G. A. Weiss, and J. A. Wells. 2000. High copy display of large proteins on phage for functional selections. J. Mol. Biol. 296:487–495.

147. Silverman, P. M. 1997. Towards a structural biology of bacterial conjugation. Mol. Microbiol. 23:423–429.

148. Skinner, M. M., H. Zhang, D. H. Leschnitzer, Y. Guan, H. Bellamy, R. M. Sweet, C. W. Gray, R. N. Konings, A. H. Wang, and T. C. Terwilliger. 1994. Structure of the gene V protein of bacteriophage f1 determined by multiwavelength x-ray diffraction on the selenomethionyl protein. Proc. Natl. Acad. Sci. USA 91:2071–2075.

149. Smith, G. P. 1985. Filamentous fusion phage. Novel expression vectors that display cloned antigens on the virion surface. Science 228:1315–1317.

150. Smith, G. P., and J. K. Scott. 1993. Libraries of peptides and proteins displayed on filamentous phage. Methods Enzymol. 217:228–257.
151. Spada, S., C. Krebber, and A. Pluckthun. 1997. Selectively infective phages (SIP). Biol. Chem. 378:445–456.
152. Specthrie, L., E. Bullitt, K. Horiuchi, P. Model, M. Russel, and L. Makowski. 1992. Construction of a microphage variant of filamentous bacteriophage. J. Mol. Biol. 228:720–724.
153. Stassen, A. P., E. F. Schoenmakers, M. Yu, J. G. Schoenmakers, and R. N. Konings. 1992. Nucleotide sequence of the genome of the filamentous bacteriophage I2-2: module evolution of the filamentous phage genome. J. Mol. Evol. 34:141–152.
154. Stathopoulos, C., D. R. Hendrixson, D. G. Thanassi, S. J. Hultgren, J. W. St Geme, 3rd, and R. Curtiss, 3rd. 2000. Secretion of virulence determinants by the general secretory pathway in gram-negative pathogens: an evolving story. Microbes Infect. 2:1061–1072.
155. Stengele, I., P. Bross, X. Garces, J. Giray, and I. Rasched. 1990. Dissection of functional domains in phage fd adsorption protein. Discrimination between attachment and penetration sites. J. Mol. Biol. 212:143–149.
156. Stump, M. D., A. S. Madison, R. J. Kokoska, and D. A. Steege. 1997. Filamentous phage IKe mRNAs conserve form and function despite divergence in regulatory elements. J. Mol. Biol. 266:51–65.
157. Stump, M. D., and D. A. Steege. 1996. Functional analysis of filamentous phage f1 mRNA processing sites. RNA 2:1286–1294.
158. Sun, T.-P., and R. E. Webster. 1986. *fii*, a bacterial locus required for filamentous phage infection and its relation to colicin-tolerant *tolA* and *tolB*. J. Bacteriol. 165:107–115.
159. Sun, T.-P., and R. E. Webster. 1987. Nucleotide sequence of a gene cluster involved in the entry of the E colicins and the single-stranded DNA of infecting filamentous phage into *Escherichia coli*. J. Bacteriol. 169:2667–2674.
160. Viti, F., F. Nilsson, S. Demartis, A. Huber, and D. Neri. 2000. Design and use of phage display libraries for the selection of antibodies and enzymes. Methods Enzymol. 326:253–275.
161. von Heijne, G., and Y. Gavel. 1988. Topogenic signals in integral membrane proteins. Eur. J. Biochem. 174:671–678.
162. Waldor, M. 1998. Bacteriophage biology and bacterial virulence. Trends Microbiol. 6:295–297.
163. Waldor, M. K., and J. J. Mekalanos. 1996. Lysogenic conversion by a filamentous phage encoding cholera toxin. Science 272:1910–1914.
164. Wall, D., and D. Kaiser. 1999. Type IV pili and cell motility. Mol. Microbiol. 32:1–10.
165. Wang, C. I., Q. Yang, and C. S. Craik. 1996. Phage display of proteases and macromolecular inhibitors. Methods Enzymol. 267:52–68.
166. Webster, R. E. 1999. Phage Display of Antibodies: CSH Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
167. Webster, R. E. 1991. The *tol* gene products and the import of macromolecules into *Escherichia coli*. Mol. Microbiol. 5:1005–1011.
168. Webster, R. E., R. A. Grant, and L. A. W. Hamilton. 1981. Orientation of the DNA in the filamentous bacteriophage f1. J. Mol. Biol. 152:357–374.
169. Weiner, L., J. L. Brissette, and P. Model. 1991. Stress-induced expression of the *Escherichia coli* phage shock protein operon is dependent on sigma 54 and modulated by positive and negative feedback mechanisms. Genes Dev. 5:1912–1923.
170. Weiner, L., J. L. Brissette, N. Ramani, and P. Model. 1995. Analysis of the proteins and *cis*-acting elements regulating the stress-induced phage shock protein operon. NAR 23:2030–2036.
171. Weiner, L., and P. Model. 1994. Role of an *Escherichia coli* stress-response operon in stationary-phase survival. Proc. Natl. Acad. Sci. USA 91:2191–2195.
172. Weiss, G. A., and S. S. Sidhu. 2000. Design and evolution of artificial M13 coat proteins. J. Mol. Biol. 300:213–219.
173. Weiss, G. A., J. A. Wells, and S. S. Sidhu. 2000. Mutational analysis of the major coat protein of M13 identifies residues that control protein display. Protein Sci. 9:647–654.
174. Wengelnik, K., C. Marie, M. Russel, and U. Bonas. 1996. Expression and localization of HrpA1, a protein of *Xanthomonas campestris* pv. *vesicatoria* essential for pathogenicity and induction of the hypersensitive reaction. J. Bacteriol. 178:1061–1069.
175. Westphal, S., L. Heins, J. Soll, and U. C. Vothknecht. 2001. *Vipp1* deletion mutant of *Synechocystis*: a connection between bacterial phage shock and thylakoid biogenesis? [See comments.] Proc. Natl. Acad. Sci. USA 98:4243–4248.
176. Widersten, M., L. O. Hansson, L. Tronstad, and B. Mannervik. 2000. Use of phage display and transition-state analogs to select enzyme variants with altered catalytic properties: glutathione transferase as an example. Methods Enzymol. 328:389–404.
177. Winter, G., A. D. Griffiths, R. E. Hawkins, and H. R. Hoogenboom. 1994. Making antibodies by phage display technology. Annu. Rev. Immunol. 12:433–455.
178. Woolford, J. L., J. S. Cashman, and R. E. Webster. 1974. f1 coat protein synthesis and altered phospholipid metabolism in f1 infected *Escherichia coli*. Virology 58:544–560.
179. Work, L. M., S. A. Nicklin, S. J. White, and A. H. Baker. 2002. Use of phage display to identify novel peptides for targeted gene therapy. Methods Enzymol. 346:157–176.
180. Wrighton, N. C., F. X. Farrell, R. Chang, A. K. Kashyap, F. P. Barbone, L. S. Mulcahy, D. L. Johnson, R. W. Barrett, L. K. Jolliffe, and W. J. Dower. 1996. Small peptides as potent mimetics of the protein hormone erythropoietin. Science 273:458–464.
181. Yip, Y. L., G. Smith, and R. L. Ward. 2001. Comparison of phage pIII, pVIII and GST as carrier proteins for peptide immunisation in Balb/c mice. Immunol. Lett. 79:197–202.
182. Yu, J. S., S. Madison-Antenucci, and D. A. Steege. 2001. Translation at higher than an optimal level interferes with coupling at an intercistronic junction. Mol. Microbiol. 42:821–834.

13

PRD1: Dissecting the Genome, Structure, and Entry

A. MARIKA GRAHN, SARAH J. BUTCHER, JAANA K. H. BAMFORD,
AND DENNIS H. BAMFORD

There are six extremely similar phage isolates, from different parts of the world, that all infect Gram-negative bacteria harboring a conjugative plasmid. One of these viruses is PRD1. Viruses with this morphology (protein capsid surrounding a membrane vesicle containing the linear double-stranded DNA genome) also infect Gram-positive hosts. These viruses have been classified together forming the *Tectiviridae* family (2). There have been two characteristics that have drawn considerable interest to these viruses. Firstly, the linear, DNA genome has inverted terminal repeats, a covalently bound terminal protein at the 5′ ends, and replicates by a protein-primed, sliding-back mechanism. Secondly, due to their membranes, these viruses are considered to be relatively simple model systems for the study of membrane biogenesis, structure, and assembly in a well-characterized *Escherichia coli* background.

As knowledge of the PRD1 structure accumulates, a third line of interest is building up. It appears that the replication mechanism, capsid architecture, major coat protein fold, and vertex structure strongly resemble those of human adenovirus (9, 11, 13, 19, 55). This has led to a proposal that these two viruses belong to the same lineage originating from a common ancestor that precedes the division of the bacterial and eukaryotic domains of life (12). This would mean that viruses are old, maybe older than the divergence of cellular life into three domains. It also appears that there are other viruses, infecting gram-positive bacteria, lower eukaryotes, and maybe archaea that are candidates for this adeno-PRD1 virus lineage. Using this structure-based comparison it has been possible to propose additional viral lineages that have members infecting hosts from different domains of life (4, 12). It should be noted that sequence comparison does not allow the identification of these functional and structural similarities due to the long evolutionary time span involved.

Previous literature includes three comprehensive reviews on PRD1 and related tectiviruses (5, 14, 47). These summarize the history and progress made, as well as the basic characteristics of these viruses. Therefore we only repeat valid information from these reviews when it is necessary, and concentrate on the progress made since 1994.

Genome

Recently the genome sequence of PRD1 was verified by automated polymerase chain reaction (PCR)-based cyclic DNA sequencing (GenBank Accession No. AY848689). Compared with the previously published sequence, which was determined at 37°C from plasmid templates (10), a few differences were observed. Due to two single-nucleotide insertions and one deletion, the sequence of gene *XI* is altered making the corresponding protein longer than previously published. The rest of the observed differences were single base changes that had no effect on the amino acid sequence. The length of the refined genome is 14,927 bp.

The genes and open reading frames (ORFs) encoded by the genome are presented in figure 13-1 and table 13-1. The nomenclature follows that given previously (10, 45): if an ORF has been shown to encode a protein it is considered to be a gene and it has been given a Roman numeral. The original ORF classification has been included in table 13-1 in order to make it easier to follow the previously published literature. The genes are organized into five operons. The early operons OE1 and OE2, which are transcribed inward from the linear genome ends, express the genome terminal protein, DNA polymerase, and two single-stranded DNA-binding proteins involved in replication. With a few exceptions, the genes for the structural components of the virion and the factors involved in assembly are organized according to their function into three late operons. Operon OL1 contains the genes for the spike-complex proteins, OL2 carries the genes for the nonstructural assembly factors,

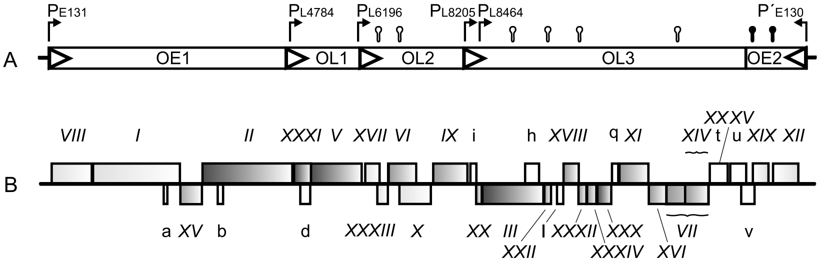


Figure 13-1 Phage PRD1 genome. A: Functional organization of the PRD1 genome including the promoters (P) and potential transcription terminators (open lollipop, forward direction; filled lollipop, reverse direction). The genome is organized into five operons. The early operons (OE1 and OE2) are transcribed inward from the genome ends. The three late operons (OL1–OL3) in the middle of the genome are transcribed in the same direction as OE1. B: The genes are identified by Roman numerals and ORFs by lower case letters according to (10). See thebacteriophages.org/frames_0130.htm for version of this figure.

Table 13-1 PRD1 Genes, Open Reading Frames (ORF), and Corresponding Proteins

Gene[a]	ORF[b]	Coordinates in PRD1 genome[b]	Protein	Mass (kDa)[c]	Description[d]
VIII		233..1012	P8	29.5	Genome terminal protein (N)
I		1016..2677	P1	63.3	DNA polymerase (N)
	(ORF a)	2415..2495		3.1	
XV		2679..3128	P15	17.3	Muramidase (L)
II		3128..4903	P2	63.7	Receptor binding (S)
	(ORF b)	3453..3587		5.1	
XXXI	(ORF c)	4907..5287	P31	13.7	Pentameric base of spike (S)
	(ORF d)	5103..5294		7.2	
V		5287..6309	P5	34.2	Trimeric spike protein (S)
XVII		6328..6588	P17	9.5	Assembly (A, N)
XXXIII	(ORF f)	6578..6784	P33	7.5	Assembly (A, N)
VI		6784..7284	P6	17.6	Minor capsid protein. DNA packaging (C, P)
X		7029..7640	P10	20.6	Assembly (A, N)
IX		7637..8320	P9	25.8	Minor capsid protein. DNA packaging ATPase (C, P)
	(ORF i)	8332..8460		4.5	
XX	(ORF j)	8460..8588	P20	4.7	DNA packaging (M, P)
III		8595..9782	P3	43.1	Major capsid protein (C)
	(ORF h)	9427..9681		9.2	
XXII	(ORF k)	9801..9944	P22	5.5	DNA packaging (M, P)
	(ORF l)	10044..10166		4.4	
XVIII	(ORF m)	10168..10440	P18	9.8	DNA delivery (M)
XXXII	(ORF n)	10440..10604	P32	5.4	DNA delivery (M)
XXXIV	(ORF o)	10617..10823	P34	6.7	(M)
XXX	(ORF p)	10833..11087	P30	9.0	Minor capsid protein (C)
	(ORF q)	11090..11200		4.2	
XI		11202..11825	P11	22.2	DNA delivery (M)
XVI	(ORF s)	11836..12189	P16	12.6	Infectivity (M)
VII		12190..12987	P7	27.1	DNA delivery, Transglycosylase (L, M)
XIV		12535..12987	P14	15.0	DNA delivery (M)
XXXV	(ORF t)	12984..13337	P35	12.8	Holin (L)
	(ORF u)	13390..13692		10.6	
	(ORF v)	13616..13888		10.2	
XIX		14132..13848[5]	P19	10.5	ssDNA binding protein (N)
XII		14687..14205[5]	P12	16.6	ssDNA binding protein (N)

[a] ORFs shown to code for functional proteins are classified as genes and are given Roman numerals.

[b] GenBank Accession No. AY848689.

[c] The mass does not include the initial methionine if not present in the mature protein.

[d] N, nonstructural early protein; M, integral membrane protein based on transmembrane helix prediction and location in the viral membrane; S, spike complex protein; A, assembly protein; P, packaging protein; C, capsid protein; L, lysis protein.

and OL3 is responsible for the coding of the viral capsid and membrane proteins (figure 13-1, table 13-1).

Structure

Early studies on the structure of PRD1 have focused on elucidating its architecture by analyzing the virion by dissociation, the effects of mutations, and the characterization of recombinant proteins (reviewed in 5). These approaches continue to be very useful (9, 17, 19, 27–29, 34, 51, 53, 54, 64) and have now been complemented by X-ray crystallography (7, 11, 12, 69, 70), Raman spectroscopy (3, 65, 66), mass spectrometry (33), small-angle X-ray scattering (63), antibody labeling (26) and cryo-electron microscopy and image reconstruction (13, 55, 57, 58).

Overall Organization

The PRD1 virion consists of a linear double-stranded DNA genome surrounded by an envelope containing host lipids and viral proteins. The major coat protein forms a protective shell over this membrane punctuated only by minor vertex and cementing proteins. Cryo-electron microscopy and image reconstruction of the virion (13, 57) revealed an icosahedral particle with average dimensions vertex to vertex of 69.8 nm, edge to edge of 65.5 nm, and facet to facet of 63.7 nm (figure 13-2A, B, C). There are 720 surface projections clustered into 240 capsomers occupying the hexavalent positions of a pseudo $T = 25$ lattice (figure 13-2A, B, G). The capsomers are trimers of the major coat protein, P3 (11, 13). Each facet of the capsid is formed from 12 copies of the P3 trimer.

The PRD1 vertices are occupied by flexible spikes in the virion, seen clearly only in electron micrographs (figure 13-2C), being averaged out in the icosahedral reconstructions because of their flexibility and nonstoichiometric association with the capsid (figure 13-2A, B, D). However, the base of the spike is clearly resolved as part of a star-shaped structure occupying the vertex (figure 13-2D). $P31^-$ mutants lack this structure along with the peripentonal trimers (55). Purified, recombinant P31 is a pentamer (33, 55) that has been modeled at low resolution from small-angle X-ray scattering data (63). When 5-fold symmetry was applied a very similar structure in shape and dimensions to the vertex base was revealed (compare figure 13-2D and E). The $P31^-$ mutant also fails to assemble P5 and P2, but the exact spatial arrangement of the spike-complex proteins is not known (55). P5, isolated as a trimer from the virion (8, 19), has at least two separate domains. The C-terminal domain is the trimerization domain, the N-terminal domain, separated from the C-terminus by a glycine-rich stretch, is the site of attachment to the virion. The N-terminus is homologous to P31, and can disrupt P31 pentamers in vitro (9, 19). So it is probable that P31 forms the pentameric base of the spike (55) and interacts with the P5 N-terminus. P2, dependent on the P5 C-terminus for assembly (9, 34), is the receptor binding protein (27, 46). Both recombinant P5 and P2 are elongated molecules (P2 is 15.5 nm long, P5 is 27 nm long; 63, 69) and are thus good candidates for forming the shaft of the spike. The X-ray structure of P2 (69, 70) revealed a club-shaped molecule with a pseudo β-propeller head (figure 13-2F, red domain) and a long tail formed from an extended β-sheet (figure 13-2F, yellow domain). The head is proposed to be the site of receptor binding, lying distal to the virus.

The Major Coat Protein

In 1999 the major coat protein structure was resolved (11), and was recently refined to 0.165 nm resolution (12). It revealed a hexagonally based trimer with three interlocking subunits, each with two eight-stranded β-barrels normal to the viral capsid. (figure 13-2H, I). The β-barrels have the same topology but no apparent sequence homology. Loops above the first β-barrel (V1) are more extensive, rising above those of the second barrel (V2). This is defined as a tower and corresponds to the 720 surface protrusions seen in the electron microscopic reconstructions (13; figure 13-2A, B, G, I). The trimer is stabilized by several interactions between the V1 of one subunit and the V2 of the adjacent one. The FG2 helix lies in the subunit interface (figure 13-2H) providing extensive hydrogen bonding for V1, V2, and the FG1 loop. In addition, the intertwining of the FG1 loops from adjacent monomers stabilizes the center of the trimer. A pseudo-atomic model of the capsid was built from the individual trimers by fitting them into density from electron microscopic reconstructions (11, 57). This demonstrated their neat interdigitation, which forms a closed shell over the surface of the membrane (figure 13-2G). It also implicated the N- and C-termini and the loop I1B2 in virion assembly as they all face the viral membrane (figure 13-2I). This has been supported by mutational analysis of key residues in the N-terminus which are thought to form an extended helix that reaches the membrane surface and stabilizes the virion during DNA packaging (13, 57, 58). The last few C-terminal residues also have an effect on assembly, supporting the hypothesis that these residues from adjacent trimers link together to help stabilize the center of the facets (58).

The location of additional capsid cementing proteins (glue proteins) was hypothesized by studying the energy of interaction between trimers in the capsid (57). Mutational studies and difference imaging have identified one potential glue protein, P30, probably located between facets, and another, as yet unidentified protein, that radiates out from the inside of the capsid vertex, between the bases of the peripentonal trimers (54, 57, 58).

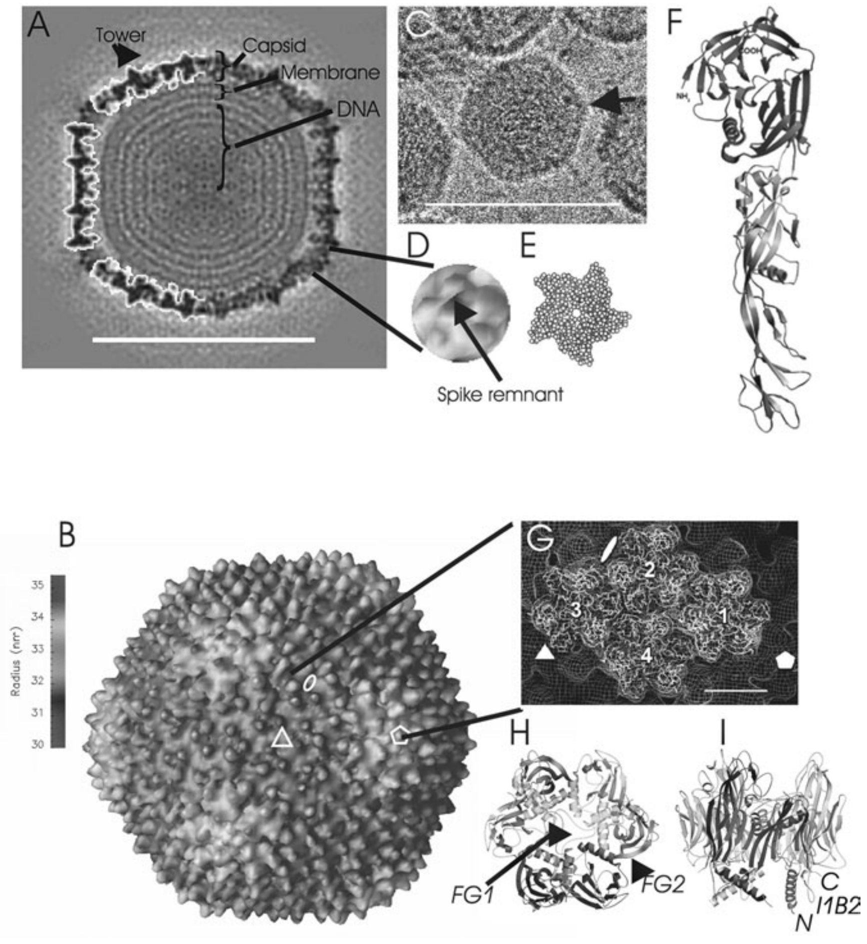


Figure 13-2 Structure of PRD1. Cryo-electron microscopy and image reconstruction of PRD1 at 1.4 nm resolution reveal the following: A: In central cross-section the organization of the virion into three layers: the outer protein capsid, the membrane bilayer, and the underlying concentric DNA rings. B: The surface representation of the icosahedral virion reconstruction viewed down a 3-fold axis of symmetry shows the pseudo $T = 25$ arrangement of the surface lattice. C: Spikes on the virion vertices that are sometimes evident in cryo-electron micrographs of the virion (arrow in C) are averaged out in the reconstruction, leaving only a small central bump as seen in the section (A) and on the surface representation (B), enlarged in (D). The vertex base (D). D: The vertex base is made of a P31 pentamer which is similar in shape and dimensions when E: modeled from SAXS data of the recombinant protein (63). F: A ribbon diagram of the atomic model of the monomeric receptor binding protein, P2 (69). The pseudo β-propeller head is the top domain, the extended β-sheet tail is the bottom domain. G: The atomic model of the trimeric major coat protein, P3, fits into four quasi-equivalent positions in the electron microscopic density (wire mesh from a 1.4 nm PRD1 $P9^-$ mutant reconstruction). The P3 C_α chains are shown fitted into the wire mesh (57). The density of P3 is also outlined in white in the left half of the section in (A). H: A ribbon diagram of the P3 trimer as seen from the outside of the virion. One monomer is highlighted to show the two adjacent β-barrels. The FG1 loops lock adjacent monomers together, the FG2 helix lies in the subunit interface. I: A side view of the P3 trimer rotated 90° relative to (H). The larger β-barrel forms all of the 720 towers seen in the reconstruction (A, B). The I1B2 loop and the N- and C-termini all face the viral membrane (12). Scale bar in (A) 50 nm, in (C) 100 nm, in (G) 5 nm. In (E) the pseudoatoms are 0.19 nm in radius. Figure (A) was prepared with SPIDER (25), (B) with opendx, (E) with rasmol (62), (F) with O (36), (F), (G) and (H) with Molscript and Raster3D (38, 44). See thebacteriophages.org/frames_0130.htm for a color version of this figure.

A small increase in α-helical content has been identified by Raman spectroscopy between a shell containing 180 trimers of P3 and 60 copies of P30, compared with trimers of P3 alone. This could be accounted for by the stabilization of the P3 subunit N- and C-terminal α-helical domains (54, 57, 58, 65).

Membrane and DNA

Laser Raman spectroscopy of PRD1 has shown that the lipids in the membrane are in the liquid crystalline phase between 5 and 50°C, and the membrane proteins are mainly α-helical (3, 10, 66). From the electron microscopic reconstructions (figure 13-2A) it is clear that the membrane is well ordered, follows the icosahedral outline of the capsid, and the leaflet centers are separated by approximately 2.9 nm. Many interactions occur between the membrane and both the capsid and the underlying DNA (13, 57, 58). The extensive interactions between the DNA and especially the phosphatidylethanolamine moieties occur without affecting the B-form backbone conformation of the DNA (66). The average separation of the concentric layers of DNA is approximately 2.5 nm, similar to that found for other bacteriophages and animal viruses (13, 57).

The structural analysis of PRD1 received a huge boost recently when diffracting crystals of a $P2^{-}$ mutant of the virion were obtained (7, 20). This ongoing structural determination will provide not only the detailed organization of the capsid but also the first high-resolution description of a native membrane as part of one of the largest structures solved to date.

Life Cycle

PRD1 is a lytic phage that exploits the transcription and possibly some of the replication functions of its host. The host cell is selected by specific recognition of a receptor on the cell surface. After adsorption, the phage genome is injected into the cell cytosol leaving the capsid outside. After production of the phage components, both virus- and host-encoded factors assist in particle assembly. Host cell lysis releases some 500 progeny virions. For a schematic presentation of the PRD1 life cycle see figure 13-3.

Receptor Recognition

PRD1 belongs to the class of broad-host-range, donor-specific phages, which infect cells only when IncP-, IncN-, or IncW-type multiple drug resistance conjugative plasmids are present. Among the hosts are several opportunistic human pathogens such as *E. coli*, *Salmonella enterica*, and *Pseudomonas aeruginosa*. Plasmid functions in the phage life cycle are dispensable after the genome has entered the cell (23, 43). These plasmids, the best studied of which is the IncP plasmid RP4 (50), encode the phage receptor on the cell surface. Receptor saturation experiments revealed approximately 25 and 60 receptors evenly distributed on the surfaces of *E. coli* and *S. enterica* cells, respectively (37). The receptor structure is accessible beyond the lipopolysaccharide layer on the cell surface since the length of the lipopolysaccharide chain does not affect PRD1 propagation (37).

The genes needed to support PRD1 entry in RP4-containing cells localize into two plasmid regions, Tra1 and Tra2, which are responsible for conjugative transfer functions (39, 68). Selection of spontaneous PRD1- resistant cells revealed that 10 Tra2 region genes (*trb*B, -C, -D, -E, -F, -G, -H, -I, -J, and -L) and one (*tra*F) from the Tra1 region are essential for PRD1 sensitivity (31). The same set of genes is the minimal requirement for RP4 self-transfer between *E. coli* cells (32). Only 1% of the spontaneous PRD1 resistant mutants were weakly transfer-proficient, suggesting that the Tra2 gene products interact with each other forming a multifunctional complex. The primary function of this apparatus is to establish an intimate cell–cell contact, the mating pair formation (Mpf), in bacterial conjugation.

The RP4 Mpf components are proposed to form a channel or a pore, facilitating DNA transfer through the membranes into the recipient cell (40). Systematic searches of the Mpf proteins for secondary structure elements, which are typical for membrane proteins, suggest that these proteins are targeted to the cell envelope. Cell fractionation studies revealed that TrbE, TrbF, TrbI, TrbL, and TraF localize to the cytoplasmic membrane and gene products containing cleavable signal peptides are exported into the periplasm (TrbC, TrbG, and TrbJ) or to the outer membrane (TrbH) independently of any other RP4 function (30). The pilin protein of RP4 (TrbC) is cyclized by an RP4-encoded serine protease, TraF, and assembled into a conjugative pilus on the cell surface (24). In the presence of all Mpf components a trans-envelope structure bridging the inner and outer membranes is formed (30); see figure 13-3. This DNA transfer complex has been shown to increase cell envelope permeability, supporting its proposed role as a conductive channel (21).

A single phage structural protein, P2, located at the vertices (figure 13-2), is responsible for PRD1 attachment to its host (46). P2 is a minor protein and its presence on the phage particle is dependent on the spike protein P5 and the penton, protein P31 (9, 55). The purified recombinant P2 is a stable monomer, which binds to its receptor with high affinity and this binding prevents phage adsorption (27). It is also a stability factor that ensures the DNA injection process does not start before the virus binds to its receptor (27).

Each of the PRD1 vertices is a metastable structure (55) and possibly capable of DNA release. The injection vertex is likely to be determined by P2 binding to the receptor (27).

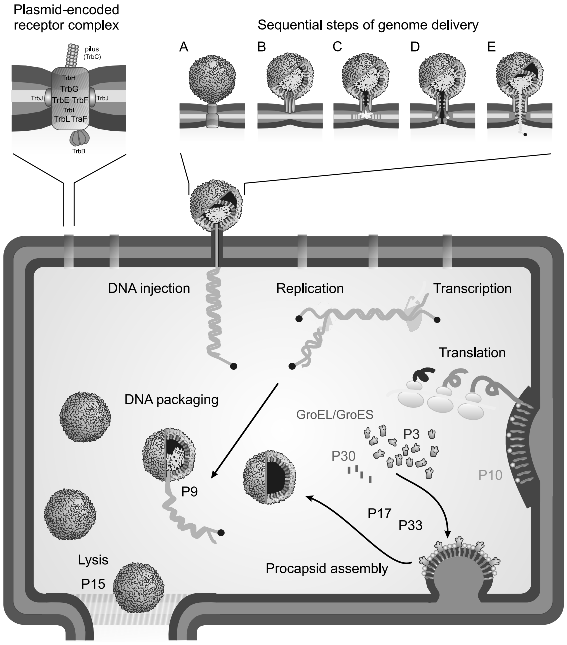


Figure 13-3 Schematic representation of the PRD1 life cycle. The vertex-associated spike complex (comprised of proteins P31, P5, and P2; see Fig. 13-2) binds to the IncP plasmid-encoded DNA transfer complex (upper left panel). The sequential steps of genome delivery are shown in the upper right panel (A–E). A: Receptor binding signals to the DNA delivery apparatus (containing at least proteins P7, P11, P14, P18, and P32) leading to considerable conformational changes in the vertex structure. B: The removal of the spike complex creates an opening in the vertex, which enables an appendage to protrude (protein P11) that penetrates the other membrane C: The lytic transglycosylase (P7) is thought to assist in genome entry by locally degrading the peptidoglycan layer D: The appendage formed extends some 35 nm (the thickness of the cell envelope), penetrating the inner membrane E: The DNA translocation in dependent on active membrane tube formation and reduction of the membrane vesicle volume assisted by at least proteins P14, P18, and P32. After DNA injection, protein-primed genome replication, transcription, and translation take place. Upon translation, the major capsid protein P3 accumulates as trimers in the host cytosol whereas those phage proteins associated with the virus membrane in the mature virion are addressed to the host cell inner membrane. Particle assembly is assisted by the host GroEL/ES complex and three phage-encoded assembly factors: the membrane-bound P10, and the soluble P17 and P33. Upon particle formation the phage-specific membrane is pinched off from the host inner membrane by the action of the scaffold protein P10 and the major coat protein P3. The minor capsid protein P30 enables the formation of an empty prohead by stabilizing the P3 interactions. The phage DNA is packaged into proheads by the packaging ATPase P9 through a unique vertex and mature virions are released upon lysis of the host cell caused by the phage-encoded muramidase P15 and holin protein P35. See thebacteriophages.org/frames_0130.htm for a color version of this figure.

The association of P2 with the receptor activates, possibly by P2 removal, the injection process. This leads to irreversible binding. Both empty and DNA-containing particles are equally tightly bound to cells, indicating that DNA injection is not a prerequisite for this tight interaction (27). Irreversible binding is not due to the receptor binding protein P2 but to an as yet unidentified component (27). After initial binding, PRD1 seems to be relatively independent in translocating the genome into the host cell cytosol (see below) and at present we do not have data about the further involvement of the IncP plasmid-encoded DNA transfer complex in the PRD1 entry process.

DNA Entry

Isolation and analysis of PRD1 mutants have resulted in the identification of eight phage-specific structural proteins essential for infectivity (9, 29, 45, 46, 51, 55). In addition to the spike-complex proteins (P2, P5, and P31) needed for adsorption, proteins P11, P14, P18, and P32 are involved in DNA delivery, since particles devoid of these proteins bind irreversibly to host cells but are noninfectious. Mutant particles missing protein P7 (a lytic transglycosylase) are infectious but the DNA entry process is delayed (51).

Efflux of cytosolic potassium has been shown to take place concomitantly with phage DNA transport (41). Electrochemical analysis of infected cells showed that PRD1 infection induces a transient efflux of K^+ and ATP from the cytosol (22). Wild-type virions also increased the permeability of the host outer membrane to lipophilic compounds, while adsorption of empty particles did not cause any changes in cell envelope permeability. Electrochemical measurements of cells infected with mutant particles missing proteins P7, P11, P14, P18, or P32 revealed that all particles, except those devoid of P11, increased the permeability of both the outer membrane and the cytoplasmic membrane, suggesting that P11 is temporally the first DNA delivery protein (28). Thus all other DNA delivery mutants are able to cause similar membrane-associated effects as the wild-type particles but DNA translocation does not occur.

It has been shown previously that the PRD1 membrane can undergo a structural transformation from a spherical vesicle to a tubular form (1, 13, 42). Empty particles with these tubes, extending from one vertex, occur with a frequency of about 10% in wild-type preparations (55). This tube formation has been suggested to be involved in PRD1 DNA injection (1, 42). Interestingly, it was observed that particles devoid of the receptor binding protein, P2, spontaneously release their DNA with concomitant formation of the tail-like membrane structure (27). Tube formation plays an important role in phage DNA transport, as absence of any of the integral viral membrane proteins P14, P18, or P32 abolishes tube formation (28, 29). Since the phenotype of the $P14^-$, $P18^-$, or $P32^-$ particles is similar and comparable to that of wild-type virions, except for the compromised tube formation, these proteins probably form a viral membrane-associated DNA translocation machinery.

Genome Replication

The genome of PRD1 is a linear double-stranded DNA molecule with proteins covalently attached to both 5′ termini and 110 bp terminal repeats (6, 59). PRD1 replicates its DNA by means of a protein-primed replication mechanism similarly to other viruses with linear double-stranded DNA genomes, including adenovirus, phage Cp-1, and the ϕ29-type phages (reviewed in 56). Development of a minimal in vitro replication system using purified components has enabled detailed characterization of the replication mechanism (61). PRD1 DNA replication starts with the formation of a covalent bond between the primer protein, P8, and the 5′ terminal nucleotide, dGMP, in a reaction catalyzed by the phage DNA polymerase, P1 (6, 60). Similarly to other DNA polymerases, P1 is activated by Mg^{2+}, but also by Mn^{2+}, which significantly stimulates the initiation reaction (16). The minimal origin of replication resides in the 20 first terminal base pairs of both genome ends and the fourth base from the 3′ end of the template directs, by base complementation, the dNMP to be linked to the terminal protein (15, 71). The 3′ end DNA sequence is maintained by a sliding-back mechanism where the polymerase complex dissociates from the template, moves back, and binds again in a stepwise fashion.

Subsequent to initiation, elongation of the initiation complex by the same DNA polymerase takes place, in a processive manner, resulting in the formation of full-length daughter DNA molecules (16, 61). The PRD1 DNA polymerase possesses, in addition to protein-primed initiation and DNA polymerization activities, a 3′–5′ exonuclease activity specific for single-stranded DNA (16, 61). This probably accounts for the proofreading capacity. Two phage-encoded DNA binding proteins, P12 and P19, are involved in replication in vivo, although they are dispensable in reactions in vitro (48, 49). Both proteins preferentially bind to single-stranded DNA, protecting it from nucleases (48, 49).

Particle Assembly

By radioactive and immunolabeling it has been shown that approximately 15 minutes after infection the major capsid protein P3 and the spike-complex proteins P2, P5, and P31 are found soluble in the host cell cytosol, whereas the phage-encoded membrane proteins (e.g., P7, P11, P14, and P18) are addressed to the host cell cytoplasmic membrane. Correct folding of the soluble proteins and the assembly of a number of viral membrane proteins are dependent on the host GroEL/ES chaperonins (35). Upon assembly, a virus-specific patch from the host cytoplasmic membrane is translocated into the forming prohead using the membrane-bound scaffolding protein P10 (54). In addition, two small phage-encoded soluble factors are indicated in the assembly process: the tetrameric protein P17 (17) and possibly P33 (7).

Recently it has been shown that the capsid of PRD1 contains, in addition to P3, a minor component P30 (about 3% of the total mass of the capsid), which has been classified as a glue protein in analogy to adenovirus glue proteins (54). This 9 kDa minor protein plays a key role in the capsid assembly process as infection with the $P30^-$ mutant results in the formation of a virus-specific membrane vesicle with only 5–10% of P3 present. The most abundant protein on the vesicle is the scaffold protein P10 (54). Thus the vesicle probably represents an early intermediate in the assembly pathway.

The correct assembly results in a prohead with a capsid enclosing the membrane rich in phage-specific proteins. The linear DNA is packaged into the prohead by the packaging ATPase P9, which is a structural component of the mature virion. In addition to P9, two small membrane proteins, P20 and P22 (46), are essential for the stable packaging of the phage DNA, indicating that the portal structure, residing in a single vertex, containing proteins P6, P9, P20, and P22 is presumably connected to the viral membrane (26, 64).

Cell Lysis

At the end of the infection cycle, the newly synthesized progeny virions are released via host cell lysis. Two genes, *XV* and *A*, involved in this step have been identified by the analysis of phage nonsense mutants (45), suggesting that a two-component, holin–endolysin system (see chapter 10) also operates in phage PRD1. The product of gene *XV*, protein P15, has been shown to be a soluble β-1,4-*N*-acetylmuramidase that effectively degrades the peptidoglycan of the Gram-negative cell causing host cell lysis (18). Recently, it was shown that this lytic enzyme is also a phage structural component associated with the membrane via protein–protein interactions (53). The PRD1 particle carries another muramidase, protein P7, which has a lytic transglycosylase activity assisting in genome entry (51). The presence of two lytic activities may reflect the broad host range of PRD1. The *Tectiviridae* family contains viruses infecting both Gram-negative and Gram-positive cells. Interestingly, in zymograms the P15 muramidase was able to degrade the peptidoglycan isolated from a Gram-positive bacterium *Micrococcus lysodeikticus*, while P7 was active only against the Gram-negative cell wall (51).

In addition to lytic enzymes, bacteriophages quite often encode helper protein factors (holins) that facilitate the access of lytic enzymes to the susceptible bond in the cell wall, and control the timing of lysis (67). Such a factor has also been reported for PRD1 (45). This holin gene was recently identified and assigned as gene *XXXV* (52).

Conclusions

The emphasis in PRD1 research is becoming more holistic. Active areas are virus–host cell interactions, detailed linkage of atomic resolution structure to function, and virus evolution. The possibility of exploring this complex system in such detail arises because of the fundamental knowledge that has already accrued.

Note added in proof: The X-ray structure of the virion and the C-terminal domains of protein PS are now available (Abrescia et al., 2004, Nature 432, 68–74; Cockburn et al., 2004, 432, 122–125; Merckel et al., 2005, Mol. cell, April issue).

Acknowledgments

Juha Huiskonen and Roman Tuma are thanked for their help in the preparation of figure 13-2. Support from the Academy of Finland to S. J. B, J. K. H. B., and D. H. B. is greatly appreciated (Finnish Centre of Excellence Programme 2000-2005).

References

1. Bamford, D., and L. Mindich. 1982. Structure of the lipid-containing bacteriophage PRD1: disruption of wild-type and nonsense mutant phage particles with guanidine hydrochloride. J. Virol. 44:1031–1038.
2. Bamford, D. H., and H.-W. Ackermann. 2000. Family *Tectiviridae*, pp. 111–116. *In* M. H. V. van Regenmortel, C. M. Fauquet, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, J. Maniloff, M. A. Mayo, D. J. McGeoch, C. R. Pringle, R. B. Wickner (eds.) Virus Taxonomy: Classification and Nomenclature of Viruses. Academic Press, San Diego.
3. Bamford, D. H., J. K. Bamford, S. A. Towse, and G. J. Thomas, Jr. 1990. Structural study of the lipid-containing bacteriophage PRD1 and its capsid and DNA components by laser Raman spectroscopy. Biochemistry 29:5982–5987.
4. Bamford, D. H., R. M. Burnett, and D. I. Stuart. 2002. Evolution of viral structure. Theor. Pop. Biol. 61:461–470.
5. Bamford, D. H., J. Caldentey, and J. K. H. Bamford. 1995. Bacteriophage PRD1: a broad host range dsDNA tectivirus with an internal membrane. Adv. Virus Res. 45:281–319.
6. Bamford, D. H., and L. Mindich. 1984. Characterization of the DNA–protein complex at the termini of the bacteriophage PRD1 genome. J. Virol. 50:309–315.
7. Bamford, J. K., J. J. Cockburn, J. Diprose, J. M. Grimes, G. Sutton, D. I. Stuart, and D. H. Bamford. 2002. Diffraction quality crystals of PRD1, a 66-MDa dsDNA virus with an internal membrane. J. Struct. Biol. 139:103–112.
8. Bamford, J. K. H., and D. H. Bamford. 1990. Capsomer proteins of bacteriophage PRD1, a bacterial virus with a membrane. Virology 177:445–451.
9. Bamford, J. K. H., and D. H. Bamford. 2000. A new mutant class, made by targeted mutagenesis, of phage PRD1 reveals that protein P5 connects the receptor binding protein to vertex. J. Virol. 74:7781–7786.
10. Bamford, J. K. H., A.-L. Hänninen, T. M. Pakula, P. M. Ojala, N. Kalkkinen, M. Frilander, and D. H. Bamford. 1991. Genome organization of membrane-containing bacteriophage PRD1. Virology 183:658–676.
11. Benson, S. D., J. K. H. Bamford, D. H. Bamford, and R. M. Burnett. 1999. Viral evolution revealed by bacteriophage PRD1 and human adenovirus coat protein structures. Cell 98:825–833.
12. Benson, S. D., J. K. H. Bamford, D. H. Bamford, and R. M. Burnett. 2002. The X-ray crystal structure of P3, the major coat protein of the lipid-containing bacteriophage PRD1, at 1.65 Å resolution. Acta Crystallogr. D58:39–59.
13. Butcher, S. J., D. H. Bamford, and S. D. Fuller. 1995. DNA packaging orders the membrane of bacteriophage PRD1. EMBO J. 14: 6078–6086.
14. Caldentey, J., J. K. Bamford, and D. H. Bamford. 1990. Structure and assembly of bacteriophage PRD1, and *Escherichia coli* virus with a membrane. J. Struct. Biol. 104:44–51.
15. Caldentey, J., L. Blanco, D. H. Bamford, and M. Salas. 1993. In vitro replication of bacteriophage PRD1 DNA. Characterization of the protein-primed initiation site. Nucleic Acids Res. 21: 3725–3730.
16. Caldentey, J., L. Blanco, H. Savilahti, D. H. Bamford, and M. Salas. 1992. In vitro replication of bacteriophage PRD1 DNA. Metal activation of protein-primed initiation and DNA elongation. Nucleic Acids Res. 20:3971–3976.
17. Caldentey, J., A. L. Hänninen, J. M. Holopainen, J. K. Bamford, P. K. Kinnunen, and D. H. Bamford. 1999. Purification and characterization of the assembly factor P17 of the lipid-containing bacteriophage PRD1. Eur. J. Biochem. 260:549–558.

18. Caldentey, J., A.-L. Hänninen, and D. H. Bamford. 1994. Gene *XV* of bacteriophage PRD1 encodes a lytic enzyme with muramidase activity. Eur. J. Biochem. 225:341–346.
19. Caldentey, J., R. Tuma, and D. H. Bamford. 2000. Assembly of bacteriophage PRD1 spike complex: the role of the multidomain protein P5. Biochemistry 39:10566–10573.
20. Cockburn, J. J., J. K. Bamford, J. M. Grimes, D. H. Bamford, and D. I. Stuart. 2003. Crystallization of the membrane-containing bacteriophage PRD1 in quartz capillaries by vapour diffusion. Acta Crystallogr. D Biol. Crystallogr. 59:538–540.
21. Daugelavicius, R., J. K. Bamford, A. M. Grahn, E. Lanka, and D. H. Bamford. 1997. The IncP plasmid-encoded cell envelope-associated DNA transfer complex increases cell permeability. J. Bacteriol. 179:5195–5202.
22. Daugelavicius, R., J. K. H. Bamford, and D. H. Bamford. 1997. Changes in host cell energetics in response to bacteriophage PRD1 DNA entry. J. Bacteriol. 179:5203–5210.
23. Davis, T. N., E. D. Muller, and J. E. Cronan, Jr. 1982. The virion of the lipid-containing bacteriophage PR4. Virology 120:287–306.
24. Eisenbrandt, R., M. Kalkum, E.-M. Lai, R. Lurz, C. I. Kado, and E. Lanka. 1999. Conjugative pili of IncP plasmids, and the Ti plasmid T pilus are composed of cyclic subunits. J. Biol. Chem. 274:22548–22555.
25. Frank, J., B. Shimkin, and H. Dowse. 1981. SPIDER: a modular software system for electron image processing. Ultramicroscopy 6:343–358.
26. Gowen, B., J. K. H. Bamford, D. H. Bamford, and S. D. Fuller. 2003. The tailless, icosahedral membrane virus PRD1 localizes the proteins involved in genome packaging and injection at a unique vertex. J. Virol. 77:7863–7871.
27. Grahn, A. M., J. Caldentey, J. K. H. Bamford, and D. H. Bamford. 1999. Stable packaging of phage PRD1 DNA requires adsorption protein P2, which binds to the IncP plasmid-encoded conjugative transfer complex. J. Bacteriol. 181:6689–6696.
28. Grahn, A. M., R. Daugelavicius, and D. H. Bamford. 2002. Sequential model of phage PRD1 DNA delivery: active involvement of the viral membrane. Mol. Microbiol. 46:1199–1209.
29. Grahn, A. M., R. Daugelavicius, and D. H. Bamford. 2002. A small viral membrane-associated protein P32 is involved in bacteriophage PRD1 DNA entry. J. Virol. 76:4866–4872.
30. Grahn, A. M., J. Haase, D. H. Bamford, and E. Lanka. 2000. Components of the RP4 conjugative transfer apparatus form an envelope structure bridging inner and outer membranes of donor cells: implications for related macromolecule transfer systems. J. Bacteriol. 182:1564–1574.
31. Grahn, A. M., J. Haase, E. Lanka, and D. H. Bamford. 1997. Assembly of a functional phage PRD1 receptor depends on 11 genes of the IncP plasmid mating pair formation complex. J. Bacteriol. 179:4733–4740.
32. Haase, J., R. Lurz, A. M. Grahn, D. H. Bamford, and E. Lanka. 1995. Bacterial conjugation mediated by plasmid RP4: RSF1010 mobilization, donor-specific phage propagation, and pilus production require the same Tra2 core components of a proposed DNA transport complex. J. Bacteriol. 177:4779–4791.
33. Helin, J., J. Caldentey, N. Kalkkinen, and D. H. Bamford. 1999. Analysis of the multimeric state of proteins by matrix assisted laser desorption/ionization mass spectrometry after cross-linking with glutaraldehyde. Rapid Commun. Mass Spectrom. 13:185–190.
34. Huiskonen, J. T., L. Laakkonen, M. Toropainen, M. Sarvas, D. H. Bamford, and J. K. H. Bamford. 2003. Probing the ability of the coat and vertex protein of the membrane-containing bacteriophage PRD1 to display a meningococcal epitope. Virology 310:267–279.
35. Hänninen, A.-L., D. H. Bamford, and J. K. H. Bamford. 1997. Assembly of membrane-containing bacteriophage PRD1 is dependent on GroEL and GroES. Virology 227:207–210.
36. Jones, T. A., J. Y. Zou, S. W. Cowan, and N. Kjeldgaard. 1991. Improved methods for binding protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A. 47:110–119.
37. Kotilainen, M. M., A. M. Grahn, J. K. H. Bamford, and D. H. Bamford. 1993. Binding of an *Escherichia coli* double-stranded DNA virus PRD1 to a receptor coded by an IncP-type plasmid. J. Bacteriol. 175:3089–3095.
38. Kraulis, P. J. 1991. MolScript: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. 24:946–950.
39. Lessl, M., D. Balzer, K. Weyrauch, and E. Lanka. 1993. The mating pair formation system of plasmid RP4 defined by RSF1010 mobilization and donor-specific phage propagation. J. Bacteriol. 175:6415–6425.
40. Lessl, M., and E. Lanka. 1994. Common mechanisms in bacterial conjugation and Ti-mediated T-DNA transfer to plant cells. Cell 77:321–324.
41. Letellier, L., L. Plancon, M. Bonhivers, and P. Boulanger. 1999. Phage DNA transport across membranes. Res. Microbiol. 150:499–505.
42. Lundström, K. H., D. H. Bamford, E. T. Palva, and K. Lounatmaa. 1979. Lipid-containing bacteriophage PR4: structure and life cycle. J. Gen. Virol. 43:583–592.
43. Lyra, C., H. Savilahti, and D. H. Bamford. 1991. High-frequency transfer of linear DNA containing 5′-covalently linked terminal proteins: electroporation of bacteriophage PRD1 genome into *Escherichia coli*. Mol. Gen. Genet. 228:65–69.
44. Merritt, E. A., and D. J. Bacon. 1997. Raster3D photorealistic molecular graphics. Methods Enzymol. 277:505–524.
45. Mindich, L., D. Bamford, C. Goldthwaite, M. Laverty, and G. Mackenzie. 1982. Isolation of nonsense mutants of lipid-containing bacteriophage PRD1. J. Virol. 44:1013–1020.
46. Mindich, L., D. Bamford, T. McGraw, and G. Mackenzie. 1982. Assembly of bacteriophage PRD1: particle formation with wild-type and mutant viruses. J. Virol. 44:1021–1030.
47. Mindich, L., and D. H. Bamford. 1988. Lipid-containing bacteriophages, pp. 475–520. *In* R. Calendar (ed.) The Bacteriophages, vol. 2. Plenum Publishing, New York.
48. Pakula, T. M., J. Caldentey, C. Gutiérrez, O. V. M., M. Salas, and D. H. Bamford. 1993. Overproduction, purification, and characterization of DNA-binding protein P19 of bacteriophage PRD1. Gene 126:99–104.

49. Pakula, T. M., J. Caldentey, M. Serrano, C. Gutierrez, J. M. Hermoso, M. Salas, and D. H. Bamford. 1990. Characterization of a DNA binding protein of bacteriophage PRD1 involved in DNA replication. Nucleic Acids Res. 18:6553–6557.
50. Pansegrau, W., E. Lanka, P. T. Barth, D. H. Figurski, D. G. Guiney, D. Haas, D. R. Helinski, H. Schwab, V. A. Stanisich, and C. M. Thomas. 1994. Complete nucleotide sequence of Birmingham IncPαplasmids. Compilation and comparative analysis. J. Mol. Biol. 239:623–663.
51. Rydman, P. S., and D. H. Bamford. 2000. Bacteriophage PRD1 DNA entry uses a viral membrane-associated transglycosylase activity. Mol. Microbiol. 37:356–363.
52. Rydman, P. S., and D. H. Bamford. 2003. Identification and mutational analysis of bacteriophage PRD1 holin protein P35. J. Bacteriol. 185:3795–3803.
53. Rydman, P. S., and D. H. Bamford. 2002. The lytic enzyme of bacteriophage PRD1 is associated with the viral membrane. J. Bacteriol. 184:104–110.
54. Rydman, P. S., J. K. H. Bamford, and D. H. Bamford. 2001. A minor capsid protein P30 is essential for bacteriophage PRD1 assembly. J. Mol. Biol. 313:785–795.
55. Rydman, P. S., J. Caldentey, S. J. Butcher, S. D. Fuller, T. Rutten, and D. H. Bamford. 1999. Bacteriophage PRD1 contains a labile receptor-binding structure at each vertex. J. Mol. Biol. 291:575–587.
56. Salas, M. 1991. Protein-priming of DNA replication. Annu. Rev. Biochem. 60:39–71.
57. San Martín, C., R. M. Burnett, F. de Haas, R. Heinkel, T. Rutten, S. D. Fuller, S. J. Butcher, and D. H. Bamford. 2001. Combined EM/X-ray imaging yields a quasi-atomic model of the adenovirus-related bacteriophage PRD1 and shows key capsid and membrane interactions. Structure 9:917–930.
58. San Martín, C., J. T. Huiskonen, J. K. Bamford, S. J. Butcher, S. D. Fuller, D. H. Bamford, and R. M. Burnett. 2002. Minor proteins, mobile arms and membrane-capsid interactions in bacteriophage PRD1 capsid. Nat. Struct. Biol. 9:756–763.
59. Savilahti, H., and D. H. Bamford. 1986. Linear DNA replication: inverted terminal repeats of the five closely related *Escherichia coli* bacteriophages. Gene 49:199–205.
60. Savilahti, H., J. Caldentey, and D. H. Bamford. 1989. Bacteriophage PRD1 terminal protein: expression of gene *VIII* in *Escherichia coli* and purification of the functional P8 product. Gene 85:45–51.
61. Savilahti, H., J. Caldentey, K. Lundström, J. E. Syväoja, and D. H. Bamford. 1991. Overexpression, purification and characterization of *Escherichia coli* bacteriophage PRD1 DNA polymerase. In vitro synthesis of full-length PRD1 DNA with purified proteins. J. Biol. Chem. 266:18737–18744.
62. Sayle, R. A., and E. J. Milner-White. 1995. RASMOL: biomolecular graphics for all. Trends Biochem. Sci. 20:374.
63. Sokolova, A., M. Malfois, J. Caldentey, D. I. Svergun, M. H. Koch, D. H. Bamford, and R. Tuma. 2001. Solution structure of bacteriophage PRD1 vertex complex. J. Biol. Chem. 276:46187–46195.
64. Strömsten, N., D. H. Bamford, and J. K. H. Bamford. 2003. The unique vertex of bacterial virus PRD1 is connected to the viral internal membrane. J. Virol. 77:6314–6321.
65. Tuma, R., J. K. H. Bamford, D. H. Bamford, M. P. Russell, and G. J. J. Thomas. 1996. Structure, interactions and dynamics of PRD1 virus I. Coupling of subunit folding and capsid assembly. J. Mol. Biol. 257:87–101.
66. Tuma, R., J. K. H. Bamford, D. H. Bamford, and G. J. J. Thomas. 1996. Structure, interactions and dynamics of PRD1 virus. II. Organization of the viral membrane and DNA. J. Mol. Biol. 257:102–115.
67. Wang, I. N., D. L. Smith, and R. Young. 2000. Holins: the protein clocks of bacteriophage infections. Annu. Rev. Microbiol. 54:799–825.
68. Waters, V. L., B. Strack, W. Pansegrau, E. Lanka, and D. G. Guiney. 1992. Mutational analysis of essential IncPα plasmid transfer genes *traF* and *traG* and involvement of *traF* in phage sensitivity. J. Bacteriol. 174:6666–6673.
69. Xu, L., S. D. Benson, S. J. Butcher, D. H. Bamford, and R. M. Burnett. 2003. The receptor binding protein P2 of PRD1, a virus targeting antibiotic-resistant bacteria, has a novel fold suggesting multiple functions. Structure (Camb.) 11:309–322.
70. Xu, L., S. J. Butcher, S. D. Benson, D. H. Bamford, and R. M. Burnett. 2000. Crystallization and preliminary X-ray analysis of receptor-binding protein P2 of bacteriophage PRD1. J. Struct. Biol. 131:159–163.
71. Yoo, S. K., and J. Ito. 1991. Sequence requirements for protein-primed DNA replication of bacteriophage PRD1. J. Mol. Biol. 218:779-789.

14

Lipid-Containing Bacteriophage PM2, the Type Organism of *Corticoviridae*

DENNIS H. BAMFORD
JAANA K. H. BAMFORD

Bacteriophage PM2 is the only described member of the *Corticoviridae* family (1). It was isolated from seawater collected from a polluted bay off the coast of Chile (13). The original host was a common marine bacterium *Pseudoalteromonas espejiana* BAL-31 (originally *Pseudomonas* BAL-31; 14, 18), which is the source of the DNA exonuclease BAL-31. Alternatively, *Pseudoalteromonas* sp. ER72M2 obtained from the East River, New York, by Leonard Mindich (26) can be used as a host for PM2.

PM2 is the first bacteriophage in which the presence of lipids in the virion was firmly demonstrated (7, 15). The mass of the virion ($\sim 4.5 \times 10^7$ Da) is distributed among nucleic acid (14%), lipid (14%), and protein (72%) constituents (8, 9). The sedimentation coefficient ($S_{20,w}$) of the particle is 293 S and buoyant density in sucrose and cesium chloride is 1.26 g/cm^3 and 1.28 g/cm^3, respectively (8, 26). The stability of the virion is dependent on sodium and calcium ions and the virion equilibrated in sucrose is inactive (26). The viral membrane is located internally (7, 20) and it forms, together with the phage-encoded membrane-associated proteins and the phage DNA, a lipid core particle (25). An icosahedrally ordered capsid, about 60 nm in diameter, surrounds the lipid core. Thus the overall architecture of PM2 resembles that of bacteriophage PRD1, a tectivirus, which has a membrane vesicle surrounding the linear double-stranded DNA genome and an outer icosahedral capsid of approximately ~65 nm in diameter (see chapter 13).

Life Cycle and Virion Components

PM2 binds to a cell surface receptor. In rich medium the infected cells lyse 60–70 minutes after infection releasing some 300 virions (26). The replication and particle assembly seem to take place in association with the host plasma membrane. Virus-size membrane vesicles, considered to be assembly intermediates, have been depicted lined up along the host plasma membrane (11). The viral lipids are derived from the host plasma membrane (15) and the lipid composition (approximately 64% phosphatidyl glycerol, 27% phosphatidyl ethanol amine, 8% neutral lipids, and small amounts of acyl phosphatidyl glycerol; 7, 33) deviates from that of the host bacterium (5). This reflects either a selective assembly mechanism or a specific location of assembly.

Originally four structural proteins, I–IV, were discovered in the PM2 virion (12), but a higher number was also postulated (6). Recently, the development of better purification and additional disruption methods for the virion (26) as well as information obtained from the genome sequence (30) enabled the identification of ten structural proteins. They were designated P1–P10 (table 14-1), P1–P4 corresponding to the earlier described proteins I–IV. The outer capsid of PM2 is composed of the major capsid protein P2 and the host cell attachment protein P1, which is located at the 5-fold vertices (6, 21, 31). These two proteins correspond to 40–50% of the protein mass of the virion. P1 can be quantitatively removed from the virion in a monomeric form by lowering the ionic strength (25). Freezing and thawing or removal of calcium ions releases both P1 and P2 quantitatively (25). Protein P2 is released as a trimer. It has been proposed that P2 binds calcium (31), and that calcium ions are essential in the final assembly process during PM2 infection (32). Proteins P3–P8 have a predicted transmembrane region and protein P4 has been shown to bind DNA (28). Figure 14-1 schematically depicts the architecture of the PM2 virion.

The Genome

The double-stranded circular genome of PM2 is one of the most tightly supercoiled DNA molecules known (16). For this reason it has been widely used as a substrate in a

Table 14-1 PM2 Genes, Open Reading Frames (ORFs), and Corresponding Proteins

Gene/ORF[a]	Coordinates in PM2 Genome[b]	Protein	Mass kDa[c]	Description[d]
XV	550..77[e]	P15	18.1	Transcription factor (N)
ORF b	755..663[e]		5.5	
XVI	1128..850[e]	P16	10.3	Transcription factor (N)
ORF d	1359..1583		8.5	
ORF e	1580..1822		8.9	
XII	1779..3719	P12	73.4	Replication initiation protein (N)
XIII	3716..3910	P13	7.2	Transcription factor (N)
XIV	3907..4212	P14	11.0	Transcription factor (N)
ORF h	4212..4643		15.7	
IX	4615..5271	P9	24.7	ATP binding site
VII	5406..5510	P7	3.6	(M)
II	5523..6332	P2	30.2	Major capsid protein
III	6345..6659	P3	10.8	(M)
IV	6659..6781	P4	4.4	(M)
VIII	6781..7008	P8	7.3	(M)
X	7079..7918	P10	29.0	
I	7918..8925	P1	37.5	Spike protein
V	8925..9407	P5	17.9	(M)
VI	9400..9783	P6	14.3	(M)
ORF k	9780..9941		6.0	
ORF l	9922..10077		5.7	

[a] ORFs shown to code for functional proteins are classified as genes and are given a Roman numeral.

[b] GenBank Accession No. AF155037.

[c] The mass does not include the initial methionine if not present in the mature protein.

[d] N, nonstructural protein; M, integral membrane protein based on transmembrane helix prediction and location in the viral membrane.

[e] The gene/ORF is transcribed in the opposite direction to that of the rest of the genes or ORFs.

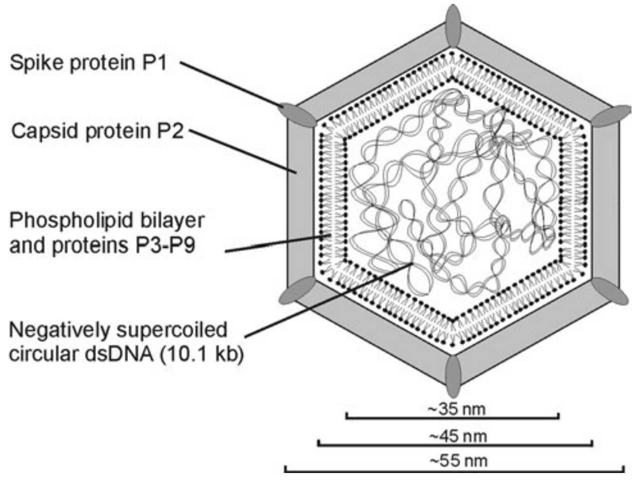

Spike protein P1
Capsid protein P2
Phospholipid bilayer and proteins P3-P9
Negatively supercoiled circular dsDNA (10.1 kb)
~35 nm
~45 nm
~55 nm

Figure 14-1 A schematic presentation of the PM2 virion architecture.

variety of enzymatic assays. The nucleotide sequence of the 10,079 bp long PM2 genome has been determined and revealed 21 putative genes (30). So far, no mutants are available for PM2 and the nomenclature for the genes and proteins is the following: a Roman numeral has been assigned to a gene when it has been confirmed to encode a protein; in other cases it is classified as an open reading frame (ORF) with a lower-case letter. The protein number (Arabic) follows the gene number with the prefix P (e.g., gene *II* encodes protein P2; table 14-1, figure 14-2; 26, 30).

Promoter mapping by primer extension has revealed putative messenger RNA start sites (29). PM2 genes are arranged in three operons—one immediate early, one early, and one late (figure 14-2)—which are expressed in a timely fashion during virus infection. Only two of these promoters, P_{1207} and P_{1193}, which guide the transcription of the leftward immediate early and the rightward early operon, respectively (figure 14-2), are functional in *E. coli* (29). However, the late operon, P_{5321} can be activated by two phage-encoded transcription factors P13 and P14 (29). Interestingly, P14 has sequence similarity to the TFIIS-type general eukaryotic transcription factors resembling most closely to those of the archaeal organisms *Thermococcus celer* (23) and *Sulfolobus acidocaldaricus* (27).

Protein P12 has also been identified on the basis of sequence similarity. It has conserved sequence motifs common to superfamily I replication initiation proteins (30). This superfamily consists of A proteins of certain bacteriophages, such as φX174 and G4, and initiation proteins of cyanobacterial and archaeal plasmids (22). Electron microscopy of PM2 DNA has revealed replication intermediates consisting of double-stranded circular molecules with growing tails no longer than the length of the genome (17). Based on this it was proposed that the replication of PM2

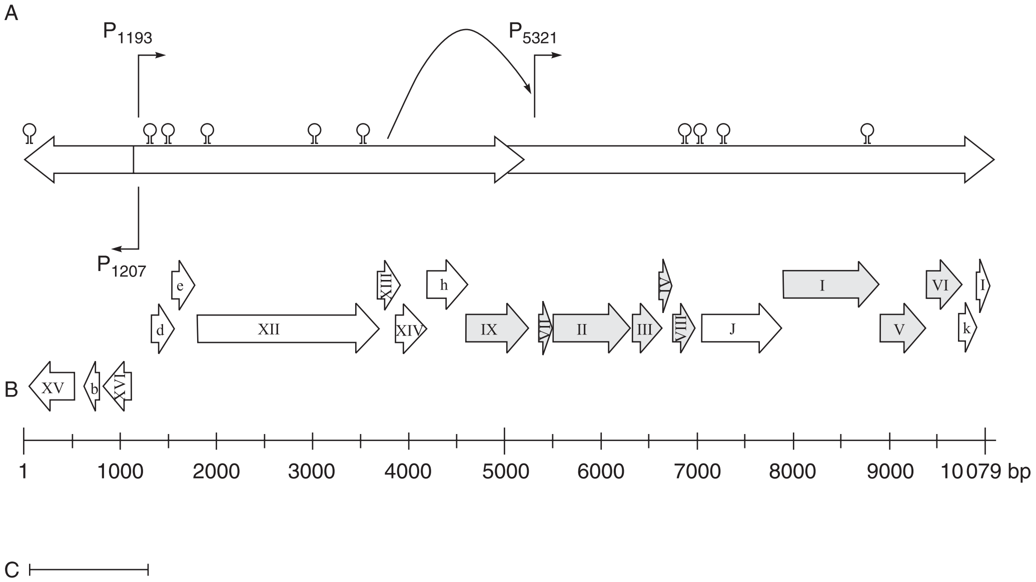


Figure 14-2 The genome of PM2 in a linear form. The circular genome is cut from a unique *Eco*RII site and the nucleotide coordinates are defined in (30). A: Promoters (P) and putative transcription terminators are presented. The genome is organized into three operons. Promoter P_{1207} is responsible for the transcription of the immediate early operon, which is transcribed in the opposite direction to the rest of the operons. The early operon is under the control of P_{1193}. The late promoter, P_{5321}, promoting the expression of the virion structural components is activated by two phage-encoded transcription factors P13 and P14. B: The genes are designated with Roman numerals and the ORFs with letters. For functions of the genome-encoded proteins see table 14-1. C: The1.2-kb-long region similar to the *Pseudoalteromonas* sp. plasmid pAS28.

DNA utilizes the rolling circle mechanism (10). The sequence similarity of protein P12 to the replication initiation proteins promoting rolling circle replication strongly supports this idea. The two early operons contain several potential genes encoding factors involved in DNA replication. This is supported by the fact that a 1.2 kb fragment from the predicted PM2 early region is homologous to the maintenance region of the *Pseudoalteromonas* plasmid pAS28 (figure 14-2; 24, 30).

PM2 has an overall architecture that resembles that of PRD1, a Tectivirus. However, the detailed capsid architecture, genome organization, and replication mechanism deviate from those of PRD1. It appears that PM2 does not belong to the same lineage of viruses as PRD1 and adenovirus (2–4). When more structural information has accumulated, it remains to be seen whether PM2 merges to a present virus lineage or forms one of its own.

Acknowledgments

This investigation was supported by research grants to J. K. H. B. and D. H. B. (Finnish Centre of Excellence Programme 2000–2005) from the Academy of Finland. Hanna Kivelä, PhD is acknowledged for providing figure 14-1 and table 14-1. Riina Hankkio, PhD is acknowledged for providing figure 14-2.

References

1. Ackermann, H.-W. 2000. Family *Corticoviridae*, pp. 117–120. *In* M. H. V. van Regenmortel, C. M. Fauquet, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, J. Maniloff, M. A. Mayo, D. J. McGeoch, C. R. Pringle, and R. B. Wickner (eds.), Virus Taxonomy: Classification and Nomenclature of Viruses. Academic Press, San Diego.
2. Bamford, D. H. 2003. Do viruses form lineages across different domains of life? Res. Microbiol. 154:231–236.
3. Bamford, D. H., R. M. Burnett, and D. I. Stuart, 2002. Evolution of viral structure. Theor. Pop. Biol. 61:461–470.
4. Benson, S. D., J. K. H. Bamford, D. H. Bamford, and R. M. Burnett. 1999. Viral evolution revealed by bacteriophage PRD1 and human adenovirus coat protein structures. Cell 98:825–833.
5. Braunstein, S. N., and R. M. Franklin. 1971. Structure and synthesis of a lipid-containing bacteriophage. V. Phospholipids of the host BAL-31 and of the bacteriophage PM2. Virology 43:685–695.
6. Brewer, G. J., and S. J. Singer. 1974. On the disposition of the proteins of the membrane-containing bacteriophage PM2. Biochemistry 13:3580–3588.
7. Camerini-Otero, R. D., and R. M. Franklin. 1972. Structure and synthesis of a lipid-containing bacteriophage. XII. The fatty acids and lipid content of bacteriophage PM2. Virology 49:385–393.

8. Camerini-Otero, R. D., and R. M. Franklin. 1975. Structure and synthesis of a lipid-containing bacteriophage. The molecular weight and other physical properties of bacteriophage PM2. Eur. J. Biochem. 53:343–348.
9. Camerini-Otero, R. D., P. N. Pusey, D. E. Koppel, D. E. Schäfer, and R. M. Franklin. 1974. Intensity fluctuation spectroscopy of laser light scattered by solutions of spherical viruses: R17, Q beta, BSV, PM2, and T7. II. Diffusion coefficients, molecular weights, solvation, and particle dimensions. Biochemistry 13:960–970.
10. Canelo, E., O. M. Phillips, and R. N. del Roure. 1985. Relating cistrons and functions in bacteriophage PM2. Virology 140:364–367.
11. Dahlberg, J. E., and R. M. Franklin. 1970. Structure and synthesis of a lipid-containing bacteriophage. IV. Electron microscopic studies of PM2-infected *Pseudomonas* BAL 31. Virology 42:1073–1086.
12. Datta, A., R. D. Camerini-Otero, S. N. Braunstein, and R. M. Franklin. 1971. Structure and synthesis of a lipid-containing bacteriophage. VII. Structural proteins of bacteriophage PM2. Virology 45:232–239.
13. Espejo, R. T., and E. S. Canelo. 1968. Properties of bacteriophage PM2: a lipid-containing bacterial virus. Virology 34:738–747.
14. Espejo, R. T. and E. S Canelo. 1968. Properties and characterization of the host bacterium of bacteriophage PM2. J. Bacteriol. 95:1887–1891.
15. Espejo, R. T., and E. S. Canelo. 1968. Origin of phospholipid in bacteriophage PM2. J. Virol. 2:1235–1240.
16. Espejo, R. T., E. S. Canelo, and R. L. Sinsheimer. 1969. DNA of bacteriophage PM2: a closed circular double-stranded molecule. Proc. Natl. Acad. of Sci. USA 63:1164–1168.
17. Espejo, R. T., E. S. Canelo, and R. L. Sinsheimer. 1971. Replication of bacteriophage PM2 deoxyribonucleic acid: a closed circular double-stranded molecule. J. Mol. Biol. 56:597–621.
18. Gauthier, G., M. Gauthier, and R. Christen. 1995. Phylogenetic analysis of the genera *Alteromonas*, *Shewanella* and *Moritella* using genes coding for small-subunit RNA sequences and division of the genus *Alteromonas* into two genera, *Alteromonas* (emended) and *Pseudoalteromonas* gen. nov., and proposal of twelve new species combinations. Int. J. Syst. Bacteriol. 45:755–761.
19. Grahn, A. M., S. J. Butcher, J. K. H. Bamford, and D. H. Bamford. 2002. PRD1: dissecting the genome, structure, and entry, pp. 00-00. *In* R. Calendar (ed.), The Bacteriophages, 2nd edn. Oxford University Press, New York.
20. Harrison, S. C., D. L. Caspar, R. D. Camerini-Otero, and R. M. Franklin. 1971. Lipid and protein arrangement in bacteriophage PM2. Nat. New Biol. 229:197–201.
21. Hinnen, R., R. Schäfer, and R. M. Franklin. 1974. Structure and synthesis of lipid-containing bacteriophage. Preparation of virus and localization of the structural proteins. Eur. J. Biochem. 50:1–14.
22. Ilyina, T. V., and E. V. Koonin. 1992. Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. Nucleic Acids Res. 20:3279–3285.
23. Kaine, B. P., I. J. Mehr, and C. R. Woese. 1994. The sequence, and its evolutionary implications, of a *Thermococcus celer* protein associated with transcription. Proc. Natl. Acad. Sci. USA 91:3854–3856.
24. Kato, J., J. Amie, J. Y. Murata, A. Kuroda, A. Mitsutani, and H. Ohtake. 1998. Development of a genetic transformation system for an alga-lysing bacterium. Appl. Environ. Microb. 64:2061–2064.
25. Kivelä, H. M., N. Kalkkinen, and D. H. Bamford. 2002. Bacteriophage PM2 has a protein capsid surrounding a spherical proteinaceous lipid core. J. Virol. 76:8169–8178.
26. Kivelä, H. M., R. H. Männistö, N. Kalkkinen, and D. H. Bamford. 1999. Purification and protein composition of PM2, the first lipid-containing bacterial virus to be isolated. Virology 262:364–374.
27. Langer, D. and W. Zillig. 1993. Putative TFIIS gene of *Sulfolobus acidocaldarius* encoding an archaeal transcription elongation factor is situated directly downstream of the gene for a small subunit of DNA-dependent RNA polymerase. Nucleic Acids Res. 21:2251.
28. Marcoli, R., V. Pirrotta, and R. M. Franklin. 1979. Interaction between bacteriophage PM2 protein IV and DNA. J. Mol. Biol. 131:107–131.
29. Männistö, R. H., A. M. Grahn, D. H. Bamford, and J. K. H. Bamford. 2003. Transcription of bacteriophage PM2 involves phage-encoded regulators of heterologous origin. J. Bacteriol. 185:3278–3287.
30. Männistö, R. H., H. M. Kivelä, L. Paulin, D. H. Bamford, and J. K. H. Bamford. 1999. The complete genome sequence of PM2, the first lipid-containing bacterial virus to be isolated. Virology 262:355–363.
31. Schäfer, R., R. Hinnen, and R. M. Franklin. 1974. Structure and synthesis of a lipid-containing bacteriophage. Properties of the structural proteins and distribution of the phospholipid. Eur. J. Biochem. 50:15–27.
32. Snipes, W., J. Cupp, J. A. Sands, A. Keith, and A. Davis. 1974. Calcium requirement for assembly of the lipid-containing bacteriophage PM2. Biochim. Biophys. Acta 339:311–322.
33. Tsukagoshi, N., M. N. Kania, and R. M. Franklin. 1976. Identification of acyl phosphatidylglycerol as a minor phospholipid of *Pseudomonas* BAL-31. Biochim. Biophys. Acta 450:131–136.

15

Single-Stranded RNA Phages

JAN VAN DUIN
NINA TSAREVA

The single-stranded RNA coliphages were discovered by Tim Loeb and Norton Zinder in 1961 as the result of a search for phages whose infection cycle depends on *E. coli* F-pili, normally used for bacterial conjugation. Loeb and Zinder plated filtered samples of raw New York City sewage on *E. coli* strains and screened for phages that would produce plaques on male (F+) but not female (F−) bacteria. The first isolate, named f1, turned out to be a filamentous phage with a single-stranded DNA genome; the second isolate, named f2, was an RNA-containing phage. Since f2 made clear plaques, Loeb and Zinder decided to concentrate their work on this phage (107).

f2 and close relatives such as MS2 and R17 represented a superb source of pure messenger RNA that could be produced in large amounts: up to 10^{13} phage particles per milliliter are made within a few hours after infection of bacterial cultures, and the phages can be easily purified. RNA phages also attracted attention because scientists were intrigued about how their RNA genomes were replicated.

Habitat

Single-stranded RNA coliphages are found wherever *E. coli* lives, for example in the intestinal tract of man and other animals. Studies have shown that sewage samples worldwide contain from 10^2 up to 10^7 RNA phage particles per milliliter (26). For humans RNA phages are harmless creatures. Several other Gram-negative bacteria can propagate their own RNA phages.

Two Major Classes of RNA Phages

The single-stranded RNA phages form the family *Leviviridae*. Based on serological cross-reactions, genetic map, and RNA size this family is divided into two genera: *Levivirus* and the *Allolevivirus* (26, 56). Leviviruses (previously called group A) have RNAs approximately 3500 nucleotides in length, and code for four proteins: maturation (or A) protein, coat protein, lysis protein, and the RNA replicase (figure 15-1). The open reading frame for the lysis protein overlaps with those of the coat and replicase proteins. MS2 phage of this genus will be used to discuss translational control in this chapter.

Phages in the *Allolevivirus* genus (group B) have a longer genome of approximately 4200 nucleotides. The difference comes mostly from a region on the genome encoding the read-through protein (previously called A1). Read-through arises when the UGA stop codon at nucleotide 1742 (Qβ) that terminates the normal coat protein gene is occasionally misread as a tryptophan codon (UGG). This occurs with a probability of about 6%. About 15 copies of the read-through protein are incorporated into the capsid (8); their precise function is unknown but the protein, together with maturation (A), is needed for the virion to be infectious. These alloleviviruses have no distinct lysis protein. Lysis is carried out by the maturation protein (37, 101). Phage Qβ has been the most intensively studied allolevivirus. Its RNA polymerase can be easily purified, while those of the leviviruses are much more difficult to work with, and so most of the work on RNA replication has been done with Qβ.

Each Genus Contains Two Species

Based on serological properties and nucleotide sequence each genus is divided into two species (26, 96). MS2 is a species I representative. Other sequenced strains within this species are f2, R17, M12, JP501 and fr* (96). Species II contains GA, KU1, JP34, and many more unsequenced

---

* Although clearly a species I phage, fr differs in many RNA structural aspects from the other strains in species I. This may relate to the fact that it was not isolated from a sewage plant but from an isolated dung-hill (by Hoffman-Berling) (C. Biebricher, personal communication).

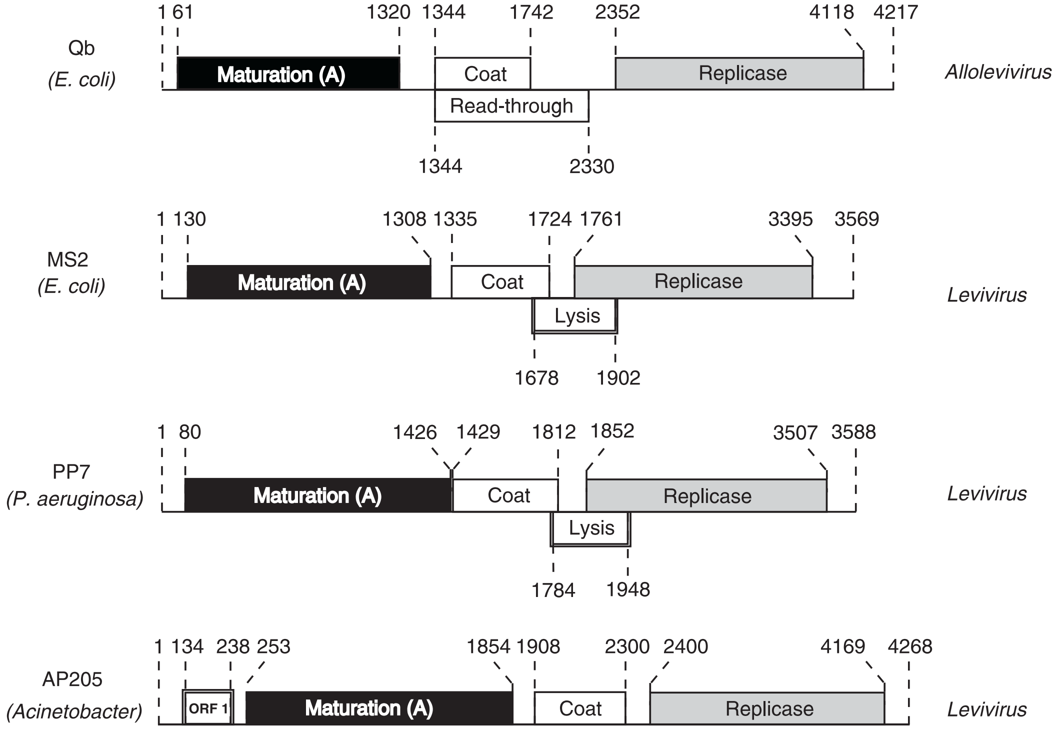

Qb
(E. coli)
1 61
1320 1344
1742
2352
4118 4217
Maturation (A)
Coat
Replicase
Read-through
1344
2330
Allolevivirus
MS2
(E. coli)
1 130
1308 1335
1724 1761
3395 3569
Maturation (A)
Coat
Replicase
Lysis
1678
1902
Levivirus
PP7
(P. aeruginosa)
1 80
1426 1429
1812 1852
3507 3588
Maturation (A)
Coat
Replicase
Lysis
1784
1948
Levivirus
AP205
(Acinetobacter)
1 134 238 253
1854 1908
2300 2400
4169 4268
ORF 1
Maturation (A)
Coat
Replicase
Levivirus

Figure 15-1 Genetic maps of four representative single-stranded RNA phages. Phages with a read-through protein are classified as alloleviviruses. ORF1 in AP 205 is thought to encode a lysis function. The map of AP205 shown here is a correction of that published in (43). The difference is due to an additional C-residue found lately at position 460.

strains. Species I has a longer and more elaborate 3′ UTR (2) and a small insertion in the replicase gene.

The alloleviviruses are also composed of two species. Qβ, ST, VK, MX1, and M11 are fully or partly sequenced (6) and belong to species III. NL95, SP, TW28, and ID2 belong to species IV (26). The RNA of this species is longer than that of III, mainly due to an insertion in the maturation protein gene.

Virion Structure

Virions contain 180 copies (90 dimers) of the coat protein arranged in a T = 3 icosahedral shell that encloses the RNA (figure 15-2A). A single copy of the maturation protein is bound to the encapsidated RNA. As mentioned above alloleviviruses contain in addition about 15 copies of the read-through protein. Encapsidated RNA is resistant to ribonuclease treatment. However, RNA in virions made with defective or missing maturation protein is sensitive to RNase. The capsid structures of Qβ, GA, PP7, and MS2 have been solved by X-ray diffraction (28, 84, 85, 88).

Natural and Artificial Infection

F-pili are the mating organelles of *E. coli* and other bacteria. They enable male bacteria to transfer a partial single-stranded copy of their chromosome to females, which do not possess pili. F-pili are made from a single protein polymerized into a long (1–2 μm), ribbon-like structure that protrudes from the cell. RNA phages do not have the specialized tail assemblies that are used by many DNA phages to inject their genomes. These phages instead subvert the F-pili to maneuver their single-stranded RNA genomes into the cell. Virions attach to the side of the pilus via their maturation protein (figure 15-2B). Upon contact with the pilus, the maturation protein is cleaved into two fragments. This releases the RNA from the virion, and it becomes sensitive to RNase degradation. How the RNA then gets inside the cell is not known. One possibility is that the pilus with the attached RNA retracts into the cell, dragging the RNA with it (63, 64).

Apart from the above natural way, there are several artificial procedures to infect the bacterial cell. In one, the outer membrane and the murein layer of the bacterium (F^+ or F^-) are removed by a controlled lysozyme treatment. The remaining spheroplast, held together by the inner membrane in isotonic medium, can be infected now with naked (+) or (−) strand RNA. If (−) strand is used, the bacterium must express replicase to obtain infection (75). It is also possible to get naked RNA into intact bacterial cells by electroporation. Alternatively, intact host cells can be transformed with a plasmid carrying the complete cDNA of an RNA phage. Transformants produce phage spontaneously (58, 83). The above procedures become important

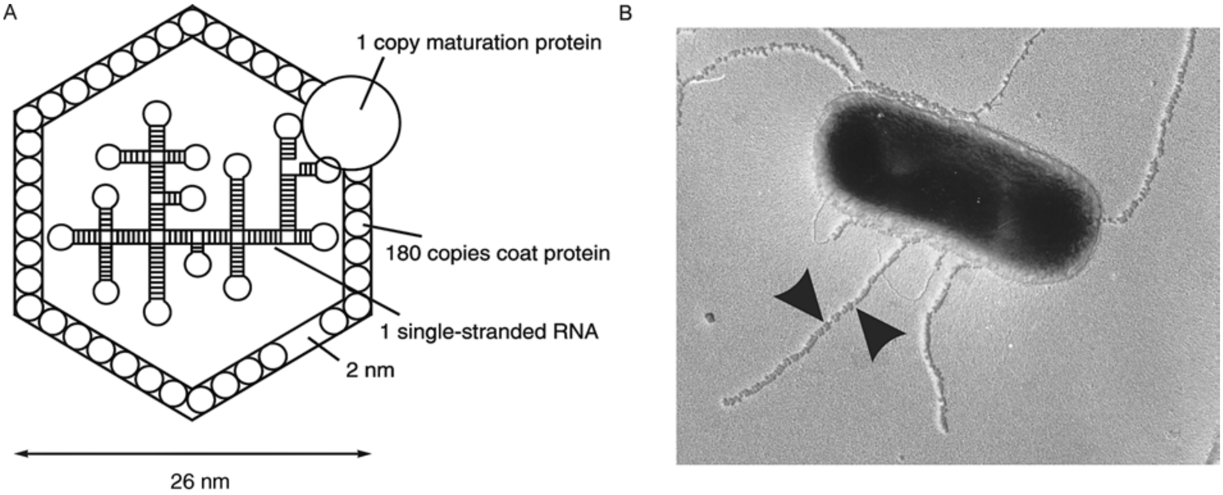


Figure 15-2 Virus representations. A: Schematic representation of a *Levivirus*. Many of the RNA hairpin loops are believed to be in contact with the coat protein dimers. The position of the maturation protein in the drawing is arbitrary. The capsid shell has a thickness of 2 nm. Alloleviviruses contain an additional approximately 15 read-through proteins in their virions. B: *Escherichia coli* bacterium with F-pili to which many MS2 phages have attached (arrows). Note that a single phage suffices for successful infection. Courtesy of A. B. Jacobson.

when one wants to infect cells with phage mutants prepared in vitro. Of the artificial methods, infection via plasmid-borne cDNA is by far the most efficient procedure.

RNA phages are among the smallest autonomous viruses known. Their task upon infecting a cell is very straightforward: make their proteins, replicate their RNA, assemble progeny phage particles, and leave the cell. However, RNA translation and RNA replication turn out to be carefully regulated, both by RNA secondary structure and by binding of coat, polymerase, and maybe maturation protein to the RNA at specific sites. Unraveling the mechanisms by which these processes are controlled has been a fascinating puzzle.

Replication Versus Translation: Competition for the Same RNA Template

Once inside the cell, phage RNA begins to function as a messenger RNA for the synthesis of phage proteins. This turns out to be a highly regulated process, for two reasons. First, different amounts of each protein are needed. For every 180 copies of the coat protein, the phage needs only 1 copy of the maturation protein to construct virions, and a few copies of the lysis protein to leave the cell. Also, since a single replicase protein can make multiple copies of RNA, far fewer replicase proteins are needed than coat proteins. Second, replication and translation of the same RNA molecule can lead to problems. Replication starts at the 3′ end of the RNA and proceeds toward the 5′ end. Translation, on the other hand, moves in the opposite sense on the RNA. If the phage had not made the proper arrangements, the polymerase and the translating ribosome would meet somewhere on the RNA and sit facing each other forever. Thus a molecule of phage RNA that begins to be translated must not be allowed to begin replication, and vice versa.

Access of ribosomes to the start sites of phage genes is strongly restricted. In fact, on intact full-length RNA only the coat gene is able to bind ribosomes directly. Translation of the lysis and replicase genes only begins once the coat gene has engaged ribosomes and is being translated. Once replicase has been made, it assembles with some host proteins (see below) to form the active polymerase (active enzyme complex of which replicase is a subunit) which wants to begin copying the same RNA molecule that is being translated. The switch from translation to replication works as follows. Although polymerase starts transcribing from the 3′ end of the phage RNA, it binds to the RNA at two internal positions, called S and M-sites. One of these, the S-site, overlaps the start of the coat gene (see figure 15-5). As a result there is competition between polymerase and ribosomes for this site. If the polymerase arrives first, there will be no new translation, allowing polymerase to copy the RNA with no oncoming traffic. On the other hand, if the ribosome binds first to the coat protein gene, polymerase will not be able to bind to its template, and therefore the ribosome is free to complete its voyage unhindered (92, 98) (see below for further details).

Mechanism of Translational Coupling Between Coat, Lysis, and Replicase Genes

How can the access of ribosomes to the start sites of the lysis and replicase genes be made dependent on translation of the coat gene? The general answer is: by changing RNA

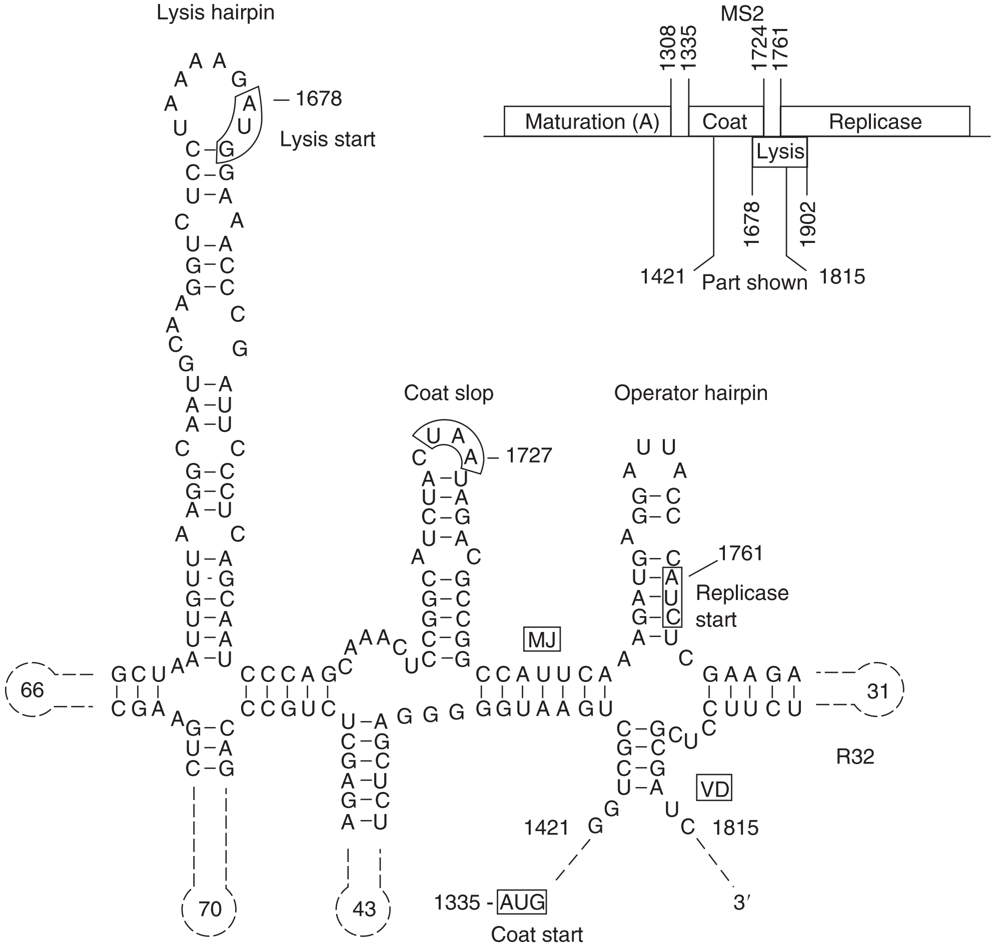


Figure 15-3 RNA secondary structure in the central part of MS2 RNA. Start codons of coat, lysis, and replicase are boxed. Numbers in dashed hairpins indicate how many nucleotides are not drawn. Structures shown are instrumental in the translational coupling between coat gene on the one hand and lysis and replicase genes on the other. MJ and VD are long-distance interactions.

secondary structures. To understand this it is useful to consider the forces that drive the binding of ribosomes to translational start regions on prokaryotic messenger RNAs. There are at least three contributors to this binding energy: (i) base complementarity between 16S ribosomal RNA and the Shine–Dalgarno sequence just upstream of the start codon on the messenger RNA; (ii) interaction of the anticodon on the initiator fmet-tRNA with the AUG start codon on the messenger RNA; and (iii) binding of ribosomal protein S1 to pyrimidine-rich sequences frequently found upstream of the Shine–Dalgarno sequence (81). If a strong pre-existing secondary structure in messenger RNA prevents binding of one or more of the above three components, there will be no ribosome binding and therefore no initiation of protein synthesis (24).

Experiments have shown that if the start codon of the coat gene is deleted or mutated, preventing ribosomes from translating this gene, neither lysis nor replicase protein is synthesized (9, 25, 69). This is because the beginning of the lysis and replicase genes lie within RNA secondary structures that are too stable to allow ribosome binding (figure 15-3).

Control of Replicase Gene Translation

Given that a ribosome contacts messenger RNA over a stretch of about nucleotides 20 upstream and 15 nt downstream of the initiator AUG, it is clear that three regions of secondary structure around the replicase start site (figure 15-3) could contribute to impeding the ribosome from binding to that site. These are first the long-distance interaction, MJ, second, the operator hairpin containing the AUG itself, and finally stem R32. When base-pairing at stem MJ is abolished by the introduction of mismatches or by deleting the sequence 1427–1434, there is a large increase in replicase synthesis, which is independent of translation of the coat gene (9, 93). Apparently, the remaining operator and R32 hairpin structures are together not strong enough to block the entry of ribosomes. The coupling then works like this. Every time a ribosome reads the coat gene it disrupts base-pairing at stem MJ. Once this happens, other ribosomes can bind to the replicase start site and initiate translation there, even though this site is some 340 nt downstream from the position where the ribosome is translating the coat protein gene. Further support for this model comes from the finding that introduction of translational stop codons into the coat protein gene upstream of nt 1427 inactivates translation of the replicase gene (the ribosome never gets to stem MJ). However, stop codons placed beyond nt 1434 allow replicase synthesis to proceed as efficiently as it does in wild-type RNA. Furthermore, the more frequently the coat-protein gene is translated, the more replicase is produced (S. H. E. van den Worm and J. van Duin, unpublished results). Coat–replicase coupling also exists in Qβ RNA, but has not been studied in great detail.

There is a second level of control of translation of the replicase gene. When the concentration of coat protein becomes sufficiently high in the cell, dimers bind to the operator hairpin, precluding further translational starts of the replicase gene by excluding ribosome binding. This protein–RNA interaction is maintained in the intact virion and has been studied exhaustively (29, 89, 90, 102, 103). It is also believed to stimulate capsid formation by serving as a nucleation site for further addition of coat dimers.

Control of Lysis Gene Translation

Independent access of ribosomes to the lysis gene start site (nt 1678) is precluded by the "lysis hairpin" (figure 15-3). In the absence of coat gene translation, the lysis gene is not expressed unless the lysis hairpin has been destabilized by disruption of base pairs (10). Surprisingly, activation of the lysis gene is not directly coupled to opening of this hairpin by ribosome movement over the overlapping coat gene (as with the replicase gene). Instead, translation starting at the lysis gene depends on termination of translation at the end of the coat gene (nt 1725). If this UAA stop codon (and the UAG stop codon that immediately follows it) are mutated so that termination does not take place until nucleotide 1749 (UGA), there is no expression of the lysis gene. On the other hand, if stop codons in the coat gene are introduced by mutagenesis between nt 1678 and 1725 or even not too far upstream of the lysis start, lysis protein is made. In fact, the closer the coat gene stop codon is placed to the lysis start codon, the more lysis protein is made (10). These results suggest that ribosomes that have reached the coat stop codon and finished translation can reinitiate at the lysis start codon after randomly drifting a short distance from their site of termination. From the ratio in which coat and lysis proteins are synthesized by phage MS2, it follows that the probability to back up to the lysis start and successfully reinitiate is only 5% in the wild-type situation.

This "scanning" model was further tested by introducing an additional start codon a short distance downstream from the authentic lysis start. Now, ribosomes used the new start codon in place of the authentic one, as it was closest to the termination triplet and thus formed a barrier that prevented ribosomes from reaching the authentic start site. If, however, coat gene termination was engineered upstream of the lysis start codon, the authentic initiation site was again used, because now this one was again closest to the termination site (1). Independent support for the scanning model comes from in vitro experiments with short messenger RNAs containing only 5 codons. Here, the ribosome shuffles back and forth between the start and stop codons, making multiple copies of a pentapeptide without ever leaving the template (65). Reinitiation is a commonly used mechanism to translate the distal reading frames in polycistronic messenger RNA. It is not known what determines the efficiency of these restarts. Ribosomes are released from the messenger RNA if no reinitiation sites are nearby.

Control of Maturation Protein Synthesis

In figure 15-4B the equilibrium secondary structure of the 5′ untranslated region (UTR) of MS2 RNA is shown. A strong stem-loop at the 5′ end is followed by a "cloverleaf structure": three hairpins enclosed by a long-range interaction, formed here by the Shine–Dalgarno (SD) sequence of the maturation gene, base-paired to an upstream complementary sequence (UCS). This long-distance interaction effectively prevents ribosome binding (30). How then is the maturation protein made? The answer is that ribosomes can only bind to the maturation start site on RNA chains that are in the process of being synthesized and have not yet reached the equilibrium folding.

Consider a growing (+) strand RNA being made on a (−) strand RNA template. By the time nucleotide 123 has emerged from the polymerase complex, all nucleotides needed to build the cloverleaf shown in figure 15-4B are present. However, translation cannot yet start because stable ribosome binding needs the start codon and sequences downstream of it, up to about nucleotide 140–145. It is known that small RNAs such as tRNA (~75 nt) fold up correctly within milliseconds (from the denatured state). Phage RNA polymerases are slow and incorporate approximately 35 nucleotides per second into a growing chain. Thus, one expects the cloverleaf structure to be formed long before the ribosome binding site is made, as addition of a further 20 nucleotides beyond nt 123 would take at least about 500 ms. However, we found that the folding of the 5′ untranslated region of MS2 RNA to its final equilibrium is unusually slow: it takes minutes to reform upon its denaturation (67). Generally speaking, such a delay can

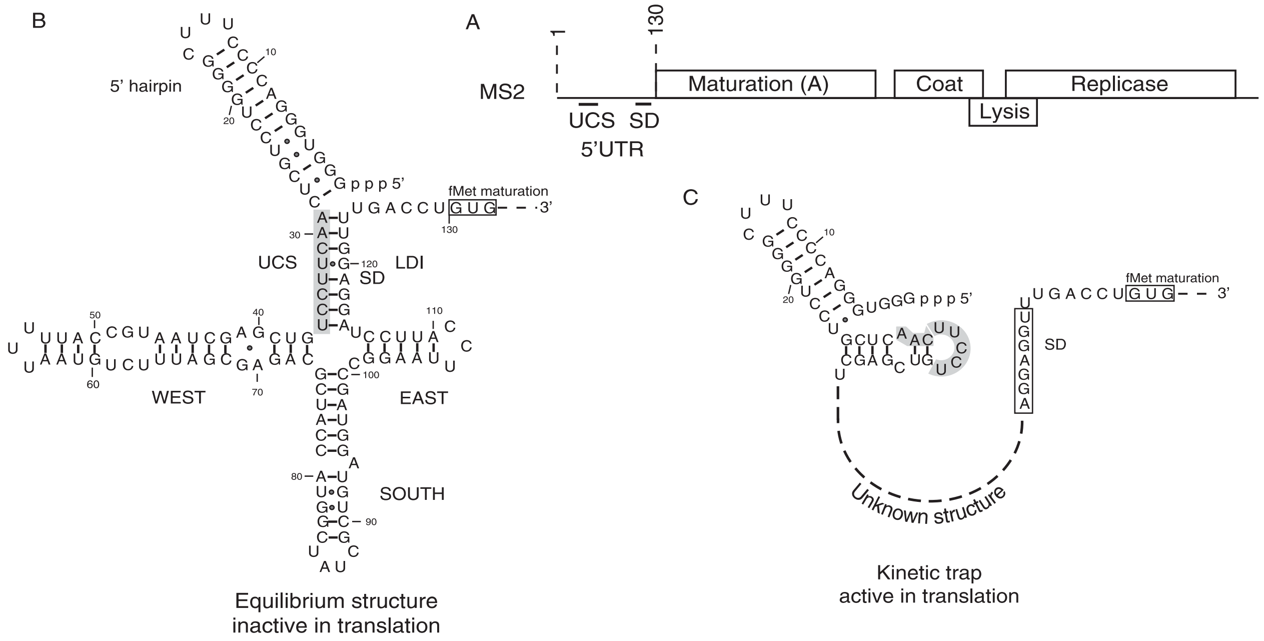


Figure 15-4 Equilibrium and nonequilibrium structure in the 5′ UTR of MS2 RNA responsible for the transient translation of the A-protein gene. A: Gene location. B: The final structure is depicted that is translationally inactive. C: The folding intermediate is shown that allows translation. Both structures are phylogenitically conserved in the *E.coli* leviviruses (95). LDI, long-distance interaction; SD, Shine–Dalgarno sequence; UCS, upstream complementary sequence. For convenience the UCS nucleotides are shaded. For simplicity, potential structure involving the start codon and surrounding has not been drawn.

only be caused by a "kinetic trap." This is an alternative folding that is stable enough to temporarily prevent the RNA from reaching the equilibrium structure.

Mutational analysis has indicated that the kinetic trap involves the small hairpin between nucleotide 25 and 43 shown in figure 15-4C (95). Formation of this hairpin excludes the inhibitory long-distance base-pairing to the SD sequence, thus permitting ribosome binding to the nonequilibrium structure. However, eventually the RNA will fold up into its equilibrium structure shown in figure 15-4B, and translation will be shut down. It is not difficult to see what the potential biological purpose of this control might be. Due to the kinetic control of maturation gene translation this gene is not accessible to ribosomes on full-length RNA. Consequently, replicase needs only to compete with ribosomes that bind at the coat start site to create a ribosome-free template, which is needed for unimpeded synthesis of (−) strand RNAs. Furthermore, it ensures that the maturation gene will only be translated once or a few times per newly synthesized (+) strand phage RNA. Indeed, only one molecule of maturation protein is needed per (+)-strand RNA molecule. In Qβ RNA, the complement to the SD sequence of the maturation protein gene lies more than 400 nt *downstream*. As a consequence the start is available at least as long as the complement has not yet been synthesized. A kinetic trap is therefore not strictly required here, but may exist to prolong the accessible state of the start site (7).

Control of Coat Gene Translation

Figure 15-5 shows the MS2 RNA and Qβ RNA structures at the start of their respective coat genes. The conspicuous feature is that, in contrast to the rest of the genome (figures 15-3, 15-4, and 15-6), there is a rather open structure formed by several single-stranded regions up- and downstream of the initiator hairpin. This whole section forms the entry site for ribosomes, polymerase, and host factor (see later). In its wild-type form the well-studied initiator hairpin of MS2 is not strong enough to downregulate translation of the coat gene: its destabilization does not increase coat protein yield. Its stabilization, on the other hand, reduces expression by a factor of 10 for every 1.4 kcal/mol stability increase (24). That is, the rate of coat gene translation is proportional to the fraction of RNA in which the initiator hairpin is in the unfolded form. One important conclusion from this finding is that ribosomes must wait until thermal motion spontaneously unfolds the hairpin before they can bind the start sequence. However, it can be calculated that this "flash exposure" lasts only about 1 μs, a time much too short for the 30S subunit to diffuse to and bind the target (25a).

A solution to this paradox is provided by "standby" binding. Here, the 30S ribosomal subunit is thought to bind, probably via protein S1, to the single-stranded regions to await the thermal unfolding. When this happens it can quickly snap in place by linear diffusion along the RNA

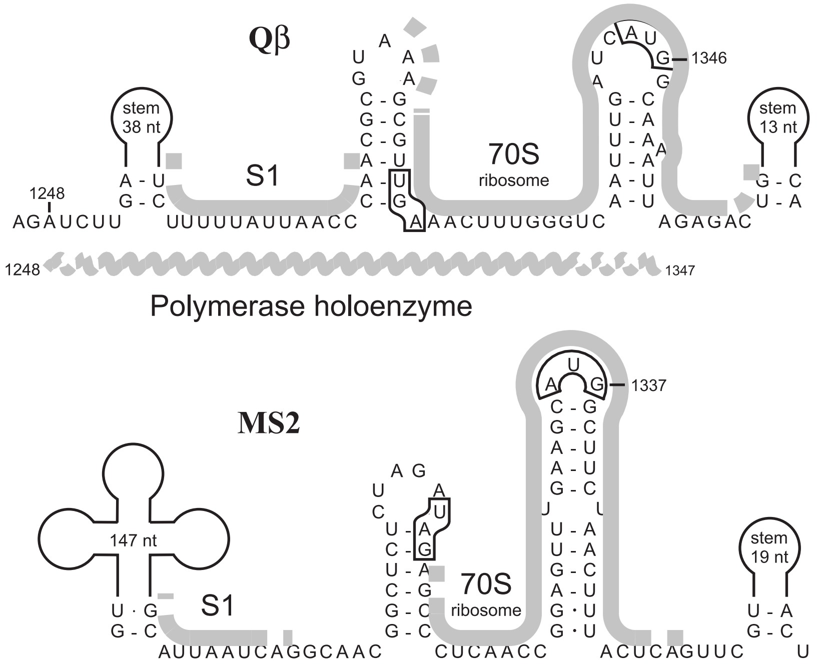


Figure 15-5 RNA structure at the coat gene start in Qβ and MS2. The stop codon of the maturation gene and start codon of the coat gene are boxed. Regions found UV-crosslinked to S1 in a 30S-phage RNA preinitiation complex are indicated (17, 17a). (In fact, the crosslink shown for MS2 was found using phage fr.) Also shown are the RNA sections protected against RNase in a 70S initiation complex (70S ribosome). The fragment of Qβ RNA protected by Qβ polymerase (the S-site) is shown as a wavy line. The exact boundaries of the binding sites are not known and this is indicated by the stippled ends.

where it will be fixed by the SD sequence and by codon–anticodon interaction of initiator-tRNA. Finally, the 50S subunit joins and the ribosome is ready to go. A noteworthy feature is that the S1 binding site seems not covered by the 70S ribosome, suggesting that S1 is only involved in early steps in initiation. The above scheme is thought to be generally valid for prokaryotic translation. It is presented here since it was developed first for the MS2 coat protein gene.

As discussed above, the only negative control on translation of the coat gene is exerted by the polymerase competing with ribosomes for their overlapping binding site.

Genome Replication Requires Four Host Cell Proteins Plus the Replicase

Once replicase is made, it can begin to generate new copies of phage RNA. The (+) stand genome is first copied into a (−) strand. This RNA is in turn used as a template to produce more (+) strand RNAs. Although fully complementary, (+) and (−) strands do not anneal to form a double-stranded RNA. Such annealing is inhibited by the high degree of internal secondary structure found in each single-strand (5, 12, 79).

When scientists isolated the RNA polymerase activity from infected cells, they found to their surprise that it is a complex made of four different proteins, only one of which (the replicase or β subunit) is coded by the phage genome (45a). The three other proteins in the complex are host proteins: ribosomal protein S1 (α subunit) and the two translation elongation factors, EF-Tu and EF-Ts (subnits γ and δ). Qβ polymerase contains EF-Tu, EF-Ts and the replicase protein in a 1:1:1 ratio. S1 is rather loosely bound and may have a stoichiometry of <1. This protein is instrumental in binding messenger RNA sequences to the ribosome during initiation of translation in bacteria (81, 91). EF-Tu positions charged tRNAs on the ribosomal A-site during elongation of growing peptide chains. This activity involves binding of GTP and its hydrolysis to GDP during the placement of the tRNA onto the ribosome. EF-Ts recycles EF-Tu·GDP to EF-Tu·GTP via the intermediate EF-Tu·EF-Ts (38).

As mentioned above, Qβ polymerase binds to two internal sites on (+) strand RNA, one of which overlaps the start site for translation of coat protein. This binding is facilitated by the S1 protein, which carries out the same function for the ribosome when it is binding to the coat protein start site. Thus the competition between ribosomes and replicase for the coat protein start site is in fact mediated by the same (cellular) protein!

Replicase initiates RNA synthesis not at the 3′ terminal A, but at the penultimate nucleotide, a C. (Terminal sequence is . . . $CCCA_{OH}$ (45a).) Thus the 5′ end of the new (−) strand

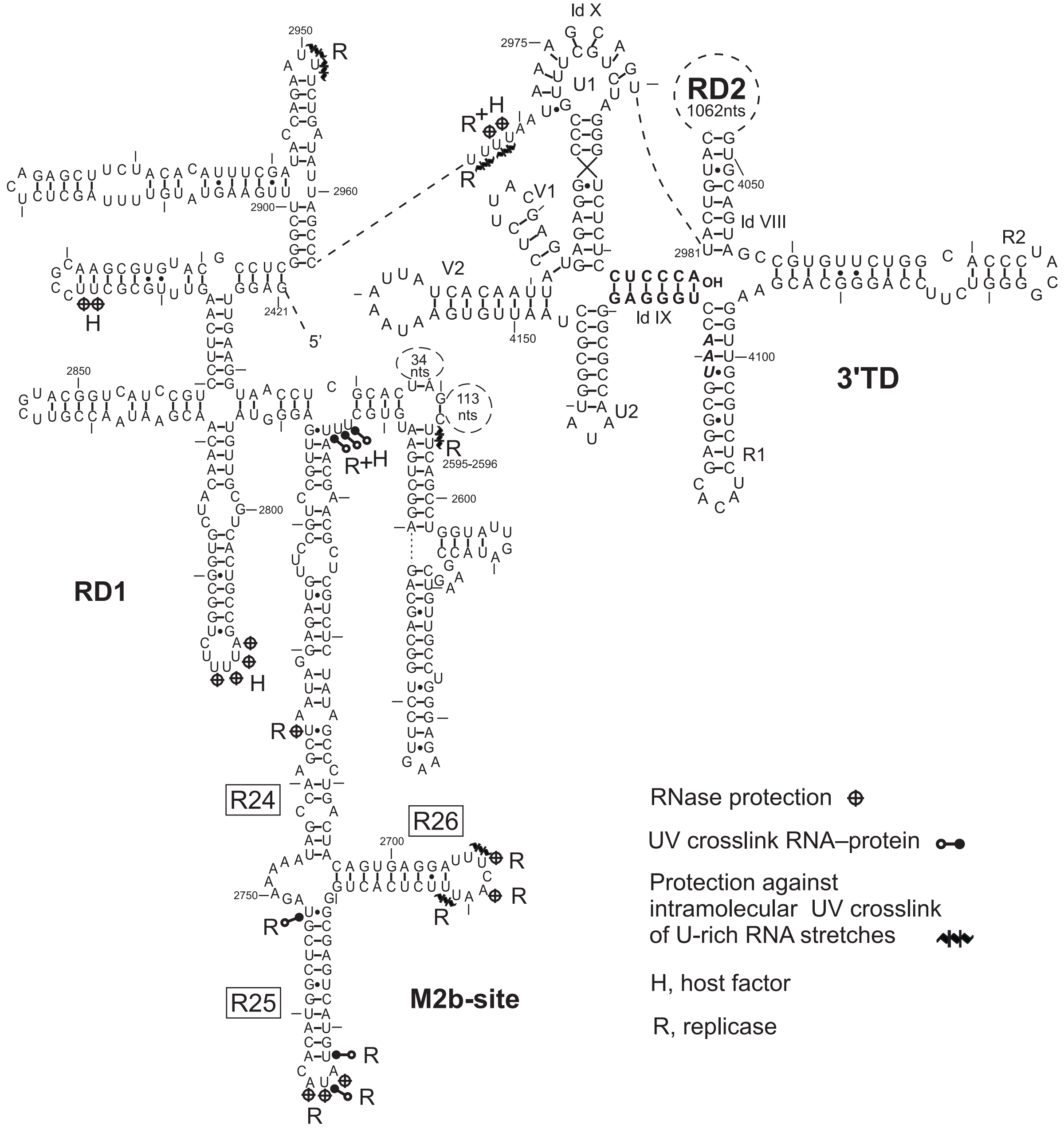

Id X
2975
2950
RD2
1062nts
U1
V1
V2
2960
2900
4050
Id VIII
2981
R2
3′TD
2421
5′
OH
Id IX
4150
34 nts
113 nts
2850
4100
U2
2595-2596
2600
R1
2800
RD1
R24
R26
2700
2750
R25
M2b-site
RNase protection ⊕
UV crosslink RNA–protein
Protection against intramolecular UV crosslink of U-rich RNA stretches
H, host factor
R, replicase
R
H
R+H

Figure 15-6 Secondary structure of S′ part of Q,β RNA. Localisation of major Qβ polymerase (R) and host factor (H) binding sites in the M-site region (replicase domain 1 or RD1) of Qβ RNA. The structure shown starts at nucleotide 2421 and runs to the very 3′ end (bold). For clarity several subdomains and RD2 are presented only as dashed circles. The region screened for protection by R and H extends from nucleotide 2550 to 3100. The folding model is based on structure probing, on phylogenetic comparison with MX1, M11, SP (33), and NL95, and on computer prediction. RD1; replicase domain 1. RD2; replicase domain 2; 3′ TD; 3′ terminal domain. The replicase stop codon at nucleotide 4119 is in bold italics (UAA). Taken from Beekwilder (6), Adhin et al. (2), and Klovins and van Duin (42).

starts with three G residues. Once synthesis has begun, it can continue in the absence of S1, EF-Tu, and EF-Ts (but the complex usually stays intact), since the replicase protein has the RNA polymerase activity. Minus strands are completed by copying the three Gs at the 5′ end of the (+) strand, and then the replicase adds on a single, untemplated A residue, so the 3′ end of the (−) strand ($CCCA_{OH}$) looks like the initial 3′ end of the (+) strand. The presence of this terminal A is slightly beneficial for transcription in the absence of host factor (see below).

Recently, it was found that Qβ polymerase must dimerize to be active as an enzyme (49). EF-Ts mutants deficient in dimerization cannot be infected by single-stranded RNA phages (37a).

Mechanism of Replication

Simple Templates

We distinguish two kinds of template: simple ones that need only the core enzyme (subunits β, γ and δ) for replication, and complex ones that need in addition protein S1 as subunit and another host factor, HF, the product of the *hfq* gene (15, 36, 45). There is only one complex template—Qβ (+) strand—but there are many simple ones, nowadays called RQ RNA because they can be used as template by Qβ replicase (22). Their size is between 50 and 250 nt. Some of them have arisen spontaneously in vitro by incubation of polymerase with high concentrations of NTPs (13). Others have been isolated from Qβ-infected cells or from cells expressing the replicase gene. They have been formed by recombination between fragments of Qβ RNA and host RNAs such as tRNA, rRNA or even RNA from phage λ (4, 53, 54, 55). Some do not even carry Qβ sequences anymore (22). Still others have been designed (106), or are SELEX products selected for binding to the polymerase core enzyme and subsequently equipped with a 3′ tail ending in $\mathrm{CCC_{OH}}$ (19). They have no biological function and can be considered as selfish RNA. Surprisingly, Qβ (−) strand is also a simple template. All RQ RNAs, including Qβ (−) strand, have in common that they begin with GGG, end with the unpaired $\mathrm{CCCA_{OH}}$ sequence and contain a binding site for Qβ polymerase core enzyme. The characteristics of the binding site are not very specific. All that seems to be needed is a stretch of weakly structured RNA, preferentially pyrimidine-rich (18–20). Even a simple 9-fold CCCA repeat forms a good template (86) and polyC has historically been used as matrix to assay column fractions for polymerase activity.

Which of the three subunits is responsible for binding the template? In a binary complex of core enzyme and template the RNA can be UV-crosslinked to the EF-Tu subunit. (In an elongating complex the crosslink shifted to the replicase subunit, in agreement with its polymerizing properties (20).) This result suggests that EF-Tu is the subunit that binds the RNA to the polymerase. Consistent with such a finding is that Qβ polymerases carrying mutant EF-Tu subunits that are progressively impaired in tRNA binding capacity are also progressively impaired in replication (48). It remains a mystery what the common denominator is in all these RQ RNAs and tRNA, the natural substrate of EF-Tu. tRNA certainly does not have weakly structured stretches. Maybe one should consider that a “weakly structured” region in an RQ RNA is flexible enough to fill (part of) the tRNA binding site of EF-Tu, which mostly involves contacts with the phosphate–sugar backbone anyway (57). Things are further complicated by the results of a fluorimetric assay which confirmed the strong affinity of polymerase for polypyrimidines but showed in addition binding to double-stranded RNA, including some tRNAs, with almost equal K_d values. Whether or not these templates were bound via EF-Tu or via another subunit was not tested (68). Also, one should realize that strong binding does not necessarily correlate with efficient replication (14).

The simplest picture for replication is that the template is anchored on the core enzyme by EF-Tu, whereafter the β subunit captures the $\mathrm{CCCA_{OH}}$ sequence (47). On short RNAs catching the terminus may be nearly instantaneous, but on longer RNA such as Qβ (−) strand such a process could take a long time and one expects the search to be guided by RNA folding, which should position the RNA terminus near the enzyme's active site. Indeed, a deletion analysis has shown that a long-distance interaction in the 3′ terminal region of Qβ (−) strand is required for transcription (70).

When the C-terminal 24 amino acids of replicase are deleted (up to Ala 565) the enzyme shows a decreased template specificity suggesting that initial contacts are not restricted to EF-Tu but may also involve the β subunit (34).

Complex Templates

Role of S1

Qβ (+) strand is the only template that requires S1 and HF for its copying (52, 99). Both proteins stimulate transcription about 10-fold. Several reasons for the special status of (+) strand come to mind. First, (+) strand is the only template serving also as messenger RNA and the dependence on S1 enables the necessary competition between ribosome and polymerase (see above). Second, the (+) strand is the genome and the infectious agent; a single molecule must be able to survive in the bacterial cytoplasm at least for as long as it takes to translate the replicase gene and produce the first (−) strand. The danger comes from the messenger RNA degrading machinery with RNase E as endonuclease and several 3′ exonucleases that turn over cellular RNAs, after these have been earmarked for decay by the addition of a poly(A) tail by poly(A) polymerase (82). Some data suggest that protection from the degradosome is the reason why the 3′ end of the (+) strand is taken up in (long-distance) base-pairing (ld IX, figure 15-6). First, extensions of the 3′ end with a few C residues, though relieving HF dependence and stimulating in vitro replication, are quickly lost in vivo. In addition, destabilizing ld IX, which also boosts in vitro replication, strongly reduces the titer in a wild-type *E. coli*, but much less so in a mutant devoid of poly(A) polymerase and polynucleotide phosphorylase (a 3′ exonuclease) (87). However, the flipside of RNase protection conferred by ld IX is that the terminal Cs are inaccessible and apparently require the help of extra proteins, S1 and HF, to be transcribed.

The following model for (+) strand transcription is proposed. As a first step polymerase holoenzyme binds the template at the S- and M-site via subunit S1. The arguments for S1 involvement are that core enzyme does not bind Qβ RNA. Furthermore, electron micrograph pictures show that complexes of Qβ RNA with either holoenzyme or with S1 alone produce the same looped structure (52). The loop was tentatively shown to represent the RNA between S- and M-sites. These two sites had previously been identified by sequencing the RNA fragments that remained bound to polymerase after digesting away unbound material by RNase. The S-site spans the region from nucleotide 1248 to 1347 and overlaps the coat gene start (figure 15-5). It is poorly structured while containing four stretches of single-stranded RNA (51).

Unlike the S-site, the M-site is highly structured (figure 15-6) and covers a large region (nucleotide 2500–3050). It is by and large the same area we now call replicase domain 1 (RD1) (6, 41). RNase protection and UV crosslinking experiments showed that contacts in the M-site are spread out in space and mainly involve the loops of hairpins and other single-stranded nucleotides (figure 15-6). The center of the interaction is considered to be at the so-called M2b-site (nucleotide 2663–2788). This region forms a branched stem-loop which is conserved in all single-stranded RNA coliphages (6). Changing base pairs in any of the stems has no effect on in vitro replication (72), consistent with the presence of many base-pair covariations between the various phage RNAs (6). Deleting R24, R25 or R26 from Qβ RNA results in a 4-fold decrease in replication. Deleting the complete M2b-site lowers replication to about 10% of the wild-type value. It is also interesting that in a binary complex between phage RNA and 30S ribosomes one of the UV crosslinks was between protein S1 and the loop of R26 (17). The others were with parts of the S-site (17a). Clearly S1 in the ribosome binds the same sites as S1 in the polymerase.

Although there is no direct evidence that interactions exist between Qβ RNA and polymerase subunits other than S1, it would be surprising if the RNA binding activity of EF-Tu were not exploited in (+) strand transcription. One argument for other contacts is that in the absence of S1 there is still a 10–20% transcription activity of the core enzyme (72). S1 thus greatly stimulates a reaction that already proceeds in its absence. We propose that S1 anchors the template on the polymerase in a standby complex. In this complex EF-Tu and probably the β-subunit are positioned at the 3′ end of the template where they must await the occasional thermal breathing of ld IX to access the C-residues. The role of S1 in translation is similar except that there the protein serves the ribosome in awaiting the thermal unfolding of the initiator hairpin. This scheme may also resolve the old paradox that a Qβ-bound polymerase inhibits ribosome binding to the coat start but a 70S initiation complex does not inhibit polymerase (45). We suppose that the S1-mediated competition is between the two long lived standby complexes. For all we know, the 70S initiation complex, in contrast to the 30S.phage RNA standby complex (51), has lost its contacts with the M-site and can thus no longer compete with the polymerase. Indeed, it would make sense to arrange competition between long-lived precursors rather than between short-lived intermediates present at the end of the initiation pathway.

Role of Host Factor

HF is a small heat-stable protein which forms hexamers. It does not associate with Qβ polymerase but acts directly on and in stoichiometry with the RNA (15, 45). Its stimulation of replication in vitro is about 5- to 10-fold and cells lacking HF can still be infected, but titers are reduced. In the uninfected cell HF is held accountable for many effects. The most basic seems that the protein promotes translation of the σ^{70} messenger by relieving inhibitory RNA structure present at the ribosomal binding site, presumably with the help of a small cellular RNA (dsrA RNA) (80, 50, 21). Like S1, HF has affinity for poorly structured RNA regions; HF with a preference for purines and S1 for pyrimidines.

Weber looked under the electron microscope at complexes formed between Qβ RNA and holoenzyme in the presence of HF and saw a two-loop structure. One loop was similar or identical to the one described above in the absence of HF; the second loop had formed between the 3′ end and the M-site (52). Formation of the second loop may reflect a previously established HF binding site in hairpin V2 (74).

Surprisingly, however, a complex between Qβ RNA and HF alone gave the same picture. This would mean that S1 and HF have the same or neighboring binding positions at the S- and M-sites. For the M-site this was confirmed; some nucleotides are protected by both polymerase and HF, while others are exclusively protected by HF or polymerase (figure 15-6). Another consequence of these multiple binding sites is that HF folds back the 3′ end to the S- and M-sites where the polymerase is bound. The 3′ end and the M-site are already close neighbors via pseudoknot ld X and we suppose that HF stabilizes this tertiary structure. Both HF and polymerase bind a stretch of six single-stranded nucleotide 5′ adjacent to ld X (nucleotides 2966–2971) (figure 15-6).

The best clue, however, about HF function stems from the pseudorevertant obtained by passaging Qβ through an HF-deficient host (71, 73). Five suppressor mutations were found that together raised transcription in the absence of HF from 10% to 60%. One of these (U2972C) changes a G·U to a G-C at the junction of the two stems of the pseudoknot (figure 15-6). It is possible that this mutation stabilizes the pseudoknot and thus the three-dimensional structure of Qβ. Note that disruptions of ld X fully abolish replication (14). The remaining four mutations are shown in figure 15-7, type B (right corner). Two of these (C4146U and A4148U) are difficult to interpret. The other two, however, replace a G-C by an A-U pair in ld IX. This change clearly destabilizes this stem, suggesting that HF somehow induces or protracts the presence of unpaired terminal nucleotides. If so, this is unlikely to be accomplished by inducing an

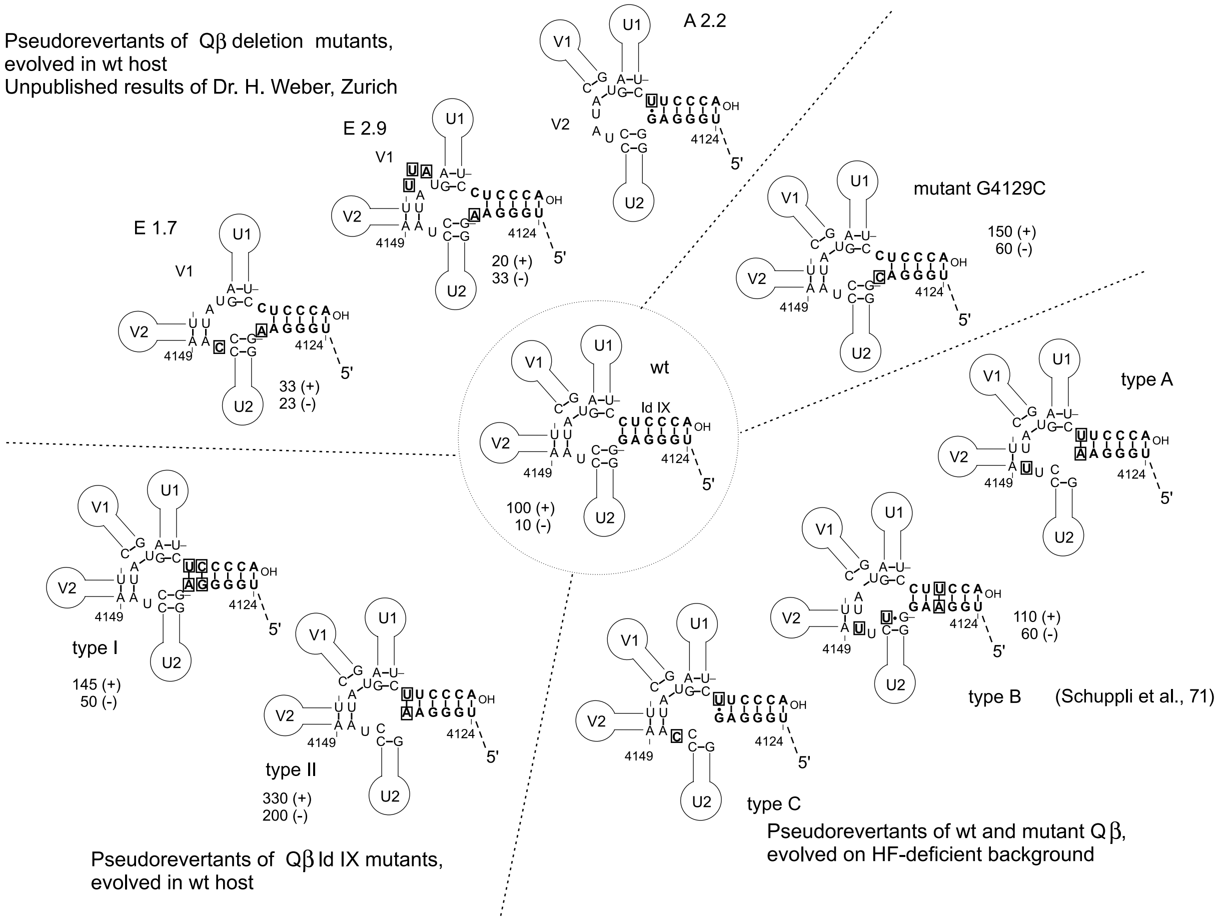

Pseudorevertants of Qβ deletion mutants,
evolved in wt host
Unpublished results of Dr. H. Weber, Zurich
A 2.2
E 2.9
E 1.7
20 (+)
33 (-)
33 (+)
23 (-)
mutant G4129C
150 (+)
60 (-)
wt
ld IX
100 (+)
10 (-)
type A
type I
145 (+)
50 (-)
type II
330 (+)
200 (-)
110 (+)
60 (-)
type B
(Schuppli et al., 71)
type C
V1
V2
U1
U2
4124
4149
5'
OH
Pseudorevertants of wt and mutant Q β,
evolved on HF-deficient background
Pseudorevertants of Qβ ld IX mutants,
evolved in wt host

Figure 15-7 Secondary structure of pseudorevertants obtained after passaging Qβ mutated in its 3′ TD through wild-type (wt) or host-factor-deficient host. The center shows the wild-type Qβ structure together with transcription efficiency in the presence (+) and absence (−) of host factor. The remarkable feature is that almost all suppressor mutations occur near the center of the five-way junction and destabilize ld IX. Furthermore, whether passaged in wild-type or host-factor-deficient hosts, all pseudorevertants show strongly decreased dependence on host factor. Mutant G4129C is a mutant, not a revertant. E1.7, E2.9, and A2.2 were sequenced only from nucleotide 2349 to the 3′ end. The transcription values refer to templates that were wild-type from nucleotide 1 to 2349 and revertant from 2349 to the 3′ end. E1.7 and E2.9 also contain the suppressor mutation (C2405U). This change is predicted to increase translation of the replicase gene as it turns an inhibitory G-C pair into a G·U pair.

alternative folding because A-U → U-A base pair reversals at positions 4128 and 4185 designed to block formation of several base-pairing alternatives had little or no effect either on HF dependence of transcription or on titer (87). Therefore, it seems that HF aids in opening ld IX. We do not know how this is achieved but one of many possibilities is by binding to the UGGGAG moiety upon spontaneous breathing of ld IX, thus slowing down the backward reaction. However, the action of HF is not simply to “melt” the RNA because the dependence on the protein *increases* with temperature.

Wild-type Qβ produces only a single HF-independent revertant. To obtain more revertants we evolved several phages with mutations in ld IX and stem-loop V1 in an HF-deficient host. Six different starting mutants produced only two additional revertants whose probable structure is shown in figure 15-7 (type A and type C). Again ld IX has been destabilized by replacement of G-C by A-U or G·U, though at a different position than in the Weber revertant (type B). In addition, G4130 is deleted and there are changes in the 4146–4148 region as observed in the original Weber revertant (73). When those six starting mutants are evolved in a wild-type host we obtain two revertants. Type I has two base pair changes in ld IX, the other (type II) is identical to type A except that it does not have the A4148U change (figure 15-7). An interesting observation is that type I and type II, even though obtained by passage through wild-type *E. coli*, show strongly decreased dependence on HF. At the same time transcription of revertant RNA is usually higher than that of wild-type. For type II, for instance, transcription in the presence and absence of HF is 330% and 200%, respectively, of the value obtained for wild-type. For the type I revertant these values are 145% and 50%. It is clear that RNA replication in wild-type is negatively controlled and that reliance on HF is not an inevitable consequence of the base-paired terminal nucleotides per se. These can efficiently be accessed in the absence of HF as the revertants show. Somehow, the dependence on HF that is incorporated in the wild-type structure makes the phage work more efficiently.

Weber and coworkers cut off one arm at a time from the five-way junction shown in figures 15-6 and 15-7. Removal of U1, U2 or V1 decreased transcription to about 4%. Deleting V2 still left 25%. Structures of revertants obtained from ΔV1 and ΔV2 are shown as E1.7, E2.9, and A2.2 (figure 15-7). (Revertant substitutions or insertions are boxed.) Again, even though passaged in a wild-type host, transcription of the revertants (E1.7 and E2.9) is almost independent of HF. For instance, for E2.9 transcription without HF is 33%, whereas with HF it is only 20%. In wild-type RNA these values are approximately 10% and 100%, respectively (D. Schuppli, J. Georgijevic and H. Weber, unpublished results). As above, the suppressor mutations destabilize ld IX at the side of the five-way junction. G-C becomes either A·C or G·U and there is again the U4147C change we have seen before. It is striking that all or almost all of the suppressor mutations localize at the five-way junction. The importance of this stucture for HF dependence is also illustrated by mutant G4129C, which shows 150% (+ HF) and 60% (−HF) transcription of the corresponding wild-type value.

Template Specificity

Polymerases of all four species have been isolated, but that of species I turned out to be unstable. All contain S1, EF-Tu, EF-Ts, and the phage-coded subunit. Species III and IV share the same host factor (HF) but MS2 and GA (32) use a different protein. These phages grow equally well in wild-type and in HF-deficient *E. coli*. The host factor for GA (GA-HF) was isolated (104). There is not a great deal of data about template specificity and the contribution of HF to specificity has not been examined. Nevertheless, some results are clear. Qβ polymerase (+ HF) will not copy MS2 or GA but it will to some extent accept species IV RNA. In contrast, GA replicase (tested without its HF) will not copy species III RNA but it will replicate species I and, surprisingly, was reported to copy SP RNA (104). Broadly speaking, replicases recognize RNA from within their genus. No experiments have been reported on whether RQ RNA can be replicated by SP or GA replicase. With regard to their simplicity one would predict they can. Neither do we know much about the template specificity of viral (−) strands.

The specificity of the polymerases for their own (+) strands is remarkable considering that the major contacts come about through the same set of sequence-nonspecific host proteins: S1, HF, and EF-Tu. It has been clear for a long time that specificity is not the result of nucleotide sequence. Rather, specificity seems to be achieved by the shape of the RNA. Only for the cognate RNA will the 3′ end be poised at the active site of the polymerase. Non-cognate RNA is predicted to be bound by the polymerase, but the 3′ end will be at an unproductive position with respect to the β subunit. The best evidence for the decisive role of RNA structure is that disruption of ld VIII or ld X in Qβ RNA leads to complete template inactivity. Pseudorevertants of disrupted ld VIII have repaired the interaction with other base pairs (41, 42).

Evolution of Man-Made Phage Mutants; Flexibility and Adaptation of the Genome

In 1978 the first infectious cDNA clone ever, that of Qβ, was constructed and shown to produce wild-type phages (83). Its potential was apparently not realized until 10 years later when infectious clones of Qβ and MS2 were prepared by Shaklee and ourselves and used for reversed genetics and other purposes. It was shown, for instance, that homologous RNA recombination does exist in RNA phages (frequency 10^{-8}) and that Qβ polymerase seems able to copy the MS2 (−) strand. That MS2 polymerase would transcribe Qβ (−) strand could not be shown (75, 76, 62).

Today infectious clones are exploited to answer a variety of questions that are difficult to solve otherwise, such as the importance of RNA structure elements for fitness. As an example we studied the role of the small hairpin containing the translational start signals for the coat gene. Accordingly, we constructed mutant 45.0 by substitutions at codon wobble positions and at upstream noncoding nucleotides (figure 15-8, center). These changes lowered the titer of the infectious clone by 4 orders of magnitude. Upon evolution this mutant produced multiple pseudorevertants that had accumulated suppressor mutations that restored the thermodynamic stability of the hairpin. As a result, the phage titer also returned to the wild-type value. This evolutionary process usually involved several subsequent base changes that stepwise improved the phenotype (58).

The conclusion from such an experiment is that the precise sequence of the coat initiator hairpin is not so important but that its stability is crucial. Separate expression experiments, already mentioned above, have shown that a hairpin more stable than wild-type dramatically decreases coat gene translation (24), and this explains the return to the wild-type stability of mutant 45.0. Surprisingly, destabilized hairpins though producing wild-type levels of coat protein also revert to wild-type stability. Possibly, the weaker initiator hairpin results in an undesired advantage for the ribosome in its competition with the polymerase.

A second example concerns the question whether or not the coupling between coat and lysis gene translation in MS2 is a coincidence resulting from the coding overlap

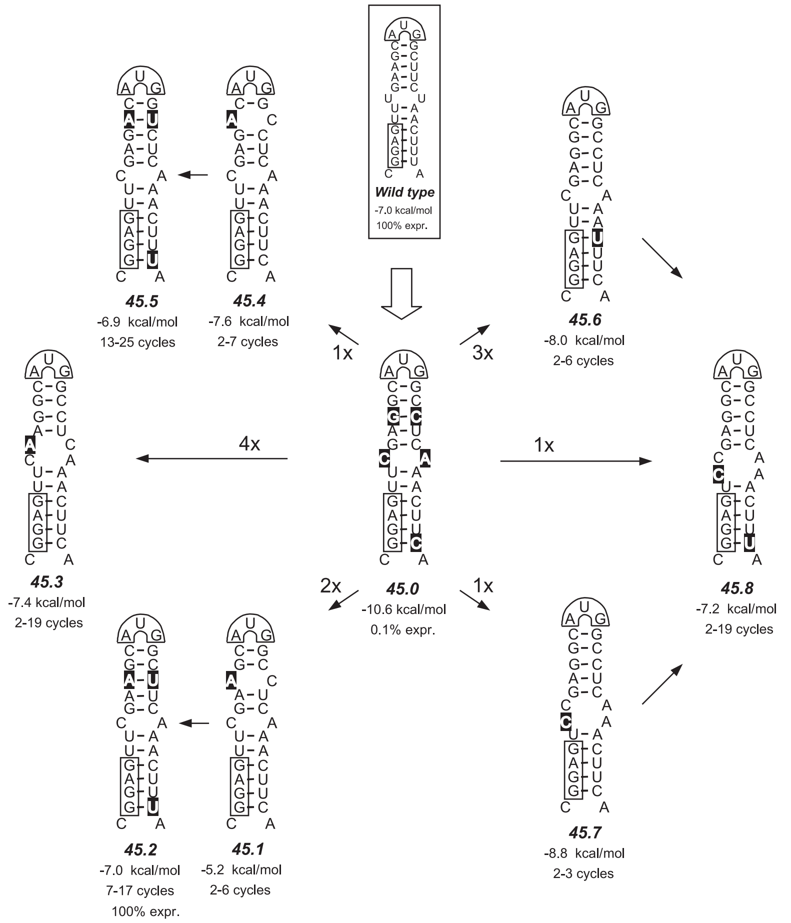


Figure 15-8 Evolution of the stabilized coat-protein initiator hairpin shown as mutant 45.0 in the center. Black boxes in 45.0 are man-made mutations. Black boxes in the pseudorevertants 45.1–45.8 are suppressor mutations selected by nature upon passaging the mutant phage. Numbers at arrows (e.g. 4×) show the number of plaques containing that specific suppressor mutation. All pseudorevertants have evolved an initiator hairpin that has a thermodynamic stability close to wild-type (boxed at top). Taken from (58).

and the prevailing RNA structure or a biological necessity. Accordingly, the lysis hairpin (figure 15-3) which is directly responsible for this coupling was mutagenized 5′ to the lysis start at neutral coat-coding positions. The mutations did not change the amount of lysis protein made, but the production was no longer dependent on coat translation. At the same time the titer of the mutant infectious clone dropped 4 logs compared with wild-type. The various pseudorevertants that were analyzed showed several second-site suppressor mutations. Their titer had again reached wild-type level and, most importantly, production of lysis was again under control of the coat gene. Therefore, the coupling is a biological necessity that adds to the fitness of the phage. Further analysis showed that control had not been regained by building a variant of the original hairpin (as in the first example). Instead, the initial man-made mutations had fully destroyed the lysis hairpin but favored several alternative foldings. These alternatives were stabilized in the pseudoreverants to the extent that independent access of ribosomes to the L start was excluded. At the same time termination-dependent reinitiation could still take place (40).

Besides showing the basics of Darwinian evolution these experiments demonstrate the importance of viral RNA structure. Without changing the coding content of the genome, the mutations in these two examples decreased the titer by 4 orders of magnitude.

More severe insults to viability are deletions, because their repair generally requires a duplication of a nearby sequence together with its adjustment to the specific needs at that position. The probability of finding such revertants in the quasi-species pool is very low and deletions are therefore often lethal. Sometimes, however, nearly incredible solutions are found to neutralize deletions. A 19 nucleotide deletion introduced in MS2 cDNA in the intercistronic region between maturation and coat genes completely removed the SD sequence of the coat as well as the initiator hairpin. Production of coat protein was undetectable.

The deletion caused a 10^{10} times decrease in titer leaving only several plaques instead of the usual 10^{11} for wild-type MS2 cDNA. These plaques showed two solutions. In one, a specific 14 nucleotide duplication near the site of the deletion restored both SD sequence and the hairpin (figure 15-9, Rev 2.1). In the other the same two features were reinstalled by a precise deletion of 6 nucleotides (figure 15-9, Rev. 1.1) that separated the two halves of a potential SD sequence (GGNNNNNNNAG). Thereafter both pseudorevertants further improved themselves by base substitutions and sometimes an additional small insertion. Most but not all of these further changes could be rationalized (legend to figure 15-9) (61).

In another example the replicase-operator hairpin of MS2 was randomized. In one of the obtained mutants, AL20, this resulted in two consecutive stop codons in the lysis reading frame (marked by a line in figure 15-10). The titer dropped 7 logs and survivors (AL20.1) had deleted the complete operator sequence (15 nucleotides) to get rid of the stop codons. In a next step six additional nucleotides were dismissed to produce the *STRANGE* pseudorevertant (46). The selection pressure for this second step is not known. Above we have seen that the operator hairpin plays important roles in shutting off replicase synthesis and in encapsidation. Its deletion shows that the shut-off of replicase is not essential but improves fitness of the phage. Furthermore, the observation that virions can still be formed shows that there must be secondary coat protein binding sites in the RNA, as suggested by Peabody on the basis of similar evolutionary experiments (66).

These evolutionary games show that there are several layers of sophistication in phages. Many interesting features such as regulation of translation, and efficient and coordinated encapsidation, are not essential for existence as a phage, but they add greatly to fitness. The generation of such elaborate mechanisms is unavoidable and is a direct consequence of the everlasting pressure to outcompete fellow sequences. In this respect it is relevant that none of the pseudorevertants ever isolated by us could successfully compete with wild-type. It is clear that in wild-type every nucleotide is carefully selected for optimal performance. This explains also why the sequence of an RNA virus is so stable in spite of the inaccuracy of the polymerase (~ 1 error per 10^3–10^4 nucleotides).

Pressures on RNA Structure by Host RNases

We have seen that RNA structure can make an active contribution to fitness, for instance by regulating translation and replication. However, one should realize that there are many other constraints that affect the final shape of the RNA. A well-known example is that large unstructured regions must be avoided as they may catalyze full duplex formation between (+) and (−) strand (5). In addition, such unstructured regions may be targets for cellular RNases.

When a noncoding hairpin in MS2 RNA was extended either with a large open loop of 26 nucleotides (figure 15-11, mutant A) or an almost fully base-paired stem (mutant C), the stem turned out to be genetically stable but the large open loop suffered deletions until an acceptable size had been reached (5–10 nucleotides; top left) (60). A similar observation was made when Qβ RNA from which a part of the read-through protein gene had been deleted was evolved in a host supplying this protein *in trans*. Plaques of these phages were initially small and titers low, but fitness improved upon passaging. Sequence analysis showed that unpaired RNA regions that resulted from the initial deletion had been removed, leading to a further abutted RNA devoid of extensive unpaired segments (3, 35).

A direct role for RNases in influencing the structure of phage RNA is not easy to prove since mutants devoid of all exonucleases or lacking endonuclease RNase E are not viable. There is, however, a null mutant for RNase III, an endonuclease active in maturation and breakdown. Its target is an uninterrupted double-stranded RNA stem of sufficient length (~ 17 bp) (23). When such a stem was incorporated in MS2 (figure 15-11, mutant R), it was genetically stable in an RNase III$^-$ mutant but not in wild-type. Here the stem evolved in many different directions, all leading to RNase III resistance (top right). Some pseudorevertants had reduced the length of the stem or had created one or more mismatches. Others showed a deletion on either side leading to big bulges. A last but most interesting category had bulges caused by the insertion of untemplated A or U residues (39). These residues turned out to be added by the host enzyme poly(A) polymerase as witnessed by the fact that such As or Us were lacking in a host devoid of this enzyme (94). In trying to reconstruct the events it was inferred that RNase III cleaves the substrate stem (either (+) or (−) strand) on one of its sides; then poly(A) polymerase adds on As to the 3′ OH created by the cleavage, thus earmarking the molecule for destruction by 3’exonucleases. Meanwhile MS2 polymerase can begin to copy such a polyadenylated or partially degraded RNA. Arriving at the site of cleavage the polymerase with the nascent chain apparently detaches and reinitiates somewhere on the other side of the cut. This may be on the added poly(A) tail or further upstream. The release and relanding of the polymerase at the physical end of the template has been observed before and has been called “run off recombination” (100).

Assembly and Release of Virions

The assembly of virions has not been studied in great detail. For MS2 it has been deduced that the first step in this process must be the binding of the maturation protein to the RNA (44). Two RNA regions are involved. One lies in the

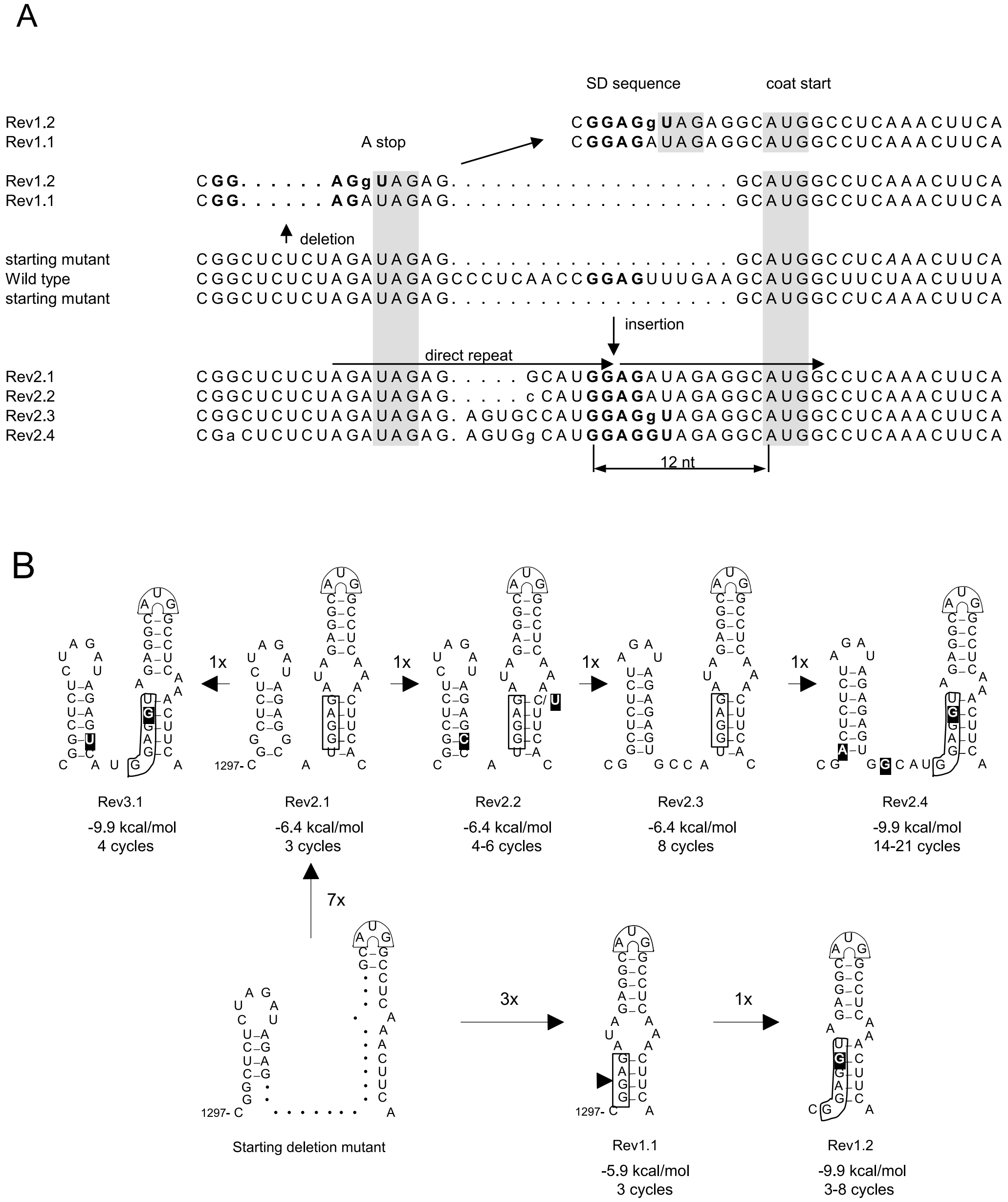

A
SD sequence
coat start
Rev1.2 CGGAGgUAGAGGCAUGGCCUCAAACUUCA
Rev1.1 CGGAGAUAGAGGCAUGGCCUCAAACUUCA
A stop
Rev1.2 CGG. . . . . . AGgUAGAG. . . . . GCAUGGCCUCAAACUUCA
Rev1.1 CGG. . . . . . AGAUAGAG. . . . . GCAUGGCCUCAAACUUCA
deletion
starting mutant CGGCUCUCUAGAUAGAG. . . . . GCAUGGCCUCAAACUUCA
Wild type CGGCUCUCUAGAUAGAGCCCUCAACCGGAGUUUGAAGCAUGGCUUCUAACUUUA
starting mutant CGGCUCUCUAGAUAGAG. . . . . GCAUGGCCUCAAACUUCA
insertion
direct repeat
Rev2.1 CGGCUCUCUAGAUAGAG. . . . . GCAUGGAGAUAGAGGCAUGGCCUCAAACUUCA
Rev2.2 CGGCUCUCUAGAUAGAG. . . . . cCAUGGAGAUAGAGGCAUGGCCUCAAACUUCA
Rev2.3 CGGCUCUCUAGAUAGAG. AGUGCCAUGGAGgUAGAGGCAUGGCCUCAAACUUCA
Rev2.4 CGaCUCUCUAGAUAGAG. AGUGgCAUGGAGGUAGAGGCAUGGCCUCAAACUUCA
12 nt
B
1x
1x
1x
1x
Rev3.1
-9.9 kcal/mol
4 cycles
Rev2.1
-6.4 kcal/mol
3 cycles
Rev2.2
-6.4 kcal/mol
4-6 cycles
Rev2.3
-6.4 kcal/mol
8 cycles
Rev2.4
-9.9 kcal/mol
14-21 cycles
7x
3x
1x
1297-
Starting deletion mutant
Rev1.1
-5.9 kcal/mol
3 cycles
Rev1.2
-9.9 kcal/mol
3-8 cycles

Figure 15-9 Evolutionary reconstruction of the deleted coat-protein initiator hairpin in MS2 RNA. A: At the sequence level and B: at the level of secondary structure. In the starting deletion mutant (bottom left) 19 nucleotides have been deleted and coat protein synthesis could no longer be detected. In Rev2.1 a specific 14 nucleotide duplication (bold italics) has occurred that restores both the initiator hairpin and the A-protein terminator hairpin as well as the Shine–Dalgarno (SD) sequence. Panel B shows that in Rev3.1 and Rev2.2 the terminator is further polished by replacing the G·G mismatch by G·U and G-C, respectively (black boxes). Rev2.3 acquires another 4 nucleotides that enlarges the distance between the two hairpins. In Rev 2.4, finally, the SD sequence is extended with one nucleotides and the distance between the hairpins is back to wild-type but for one nt. In Rev1.1, the SD sequence and the initiator hairpin are restored by a 6 nucleotides deletion. The terminator hairpin is sacrificed. In Rev1.2, the same A → G substitution is selected as in Rev2.4 and Rev3.1. This change extends the SD sequence by one nucleotide. In panel A stop codon of the A-protein gene and start codon of the coat protein are shaded. The SD sequence is in bold and suppressor mutations are in lower-case. In panel B suppressor mutations are in black boxes. SD sequences and coat-start codon are boxed. Note that the starting mutant has two accidental base substitutions, three and six nucleotide 3′ to the start codon.

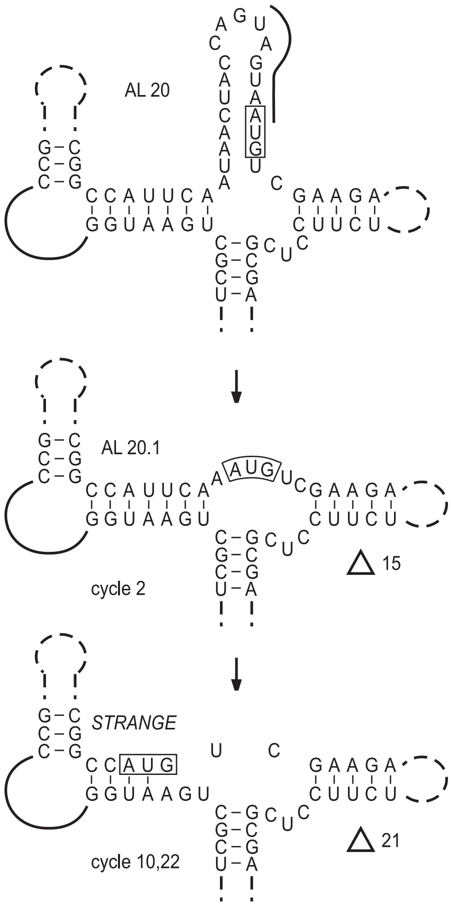


Figure 15-10 Evolution of MS2 mutant AL20. This mutant contains multiple point mutations in the replicase operator hairpin, some of which have created two consecutive stopcodons in the reading frame of the lysis gene. Evolution deletes the complete operator including the stopcodons and restores the lysis function at the expense of the operator.

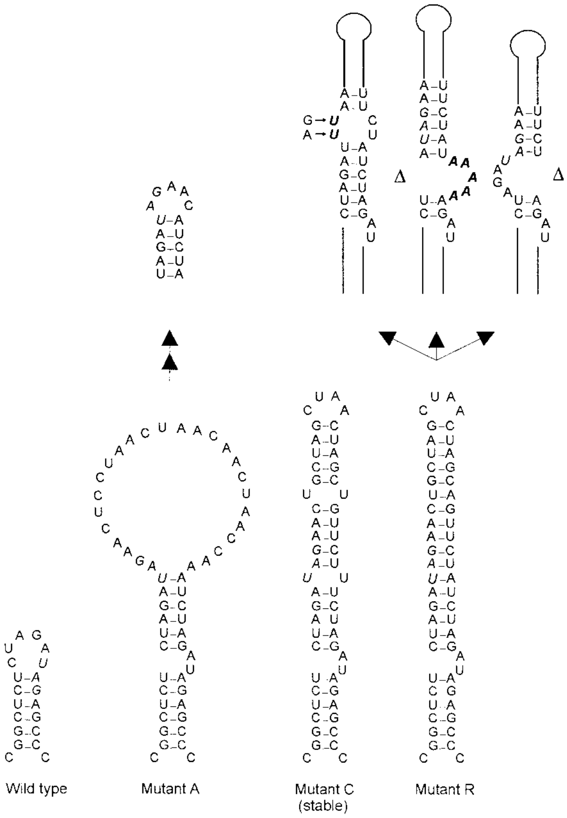


Figure 15-11 Acceptance or rejection of various RNA structure elements introduced in a noncoding region of MS2. At the left we show the wild-type structure, containing the stop codon (UAG italics) of the maturation gene. Mutant A has been given a big unstructured loop that is almost fully deleted during passaging (double arrows). Mutant C, containing a nearly perfect hairpin extension, is genetically stable for many generations. Mutant R with the perfect hairpin extension is an RNase III substrate. In RNase III-deficient strains the stem is genetically stable but in wild-type *E. coli* suppressor mutations occur that make the stem RNase III resistant (bold italics, top right). We show only some representative examples of the pseudorevertants obtained.

A-protein gene around position 400. The other involves hairpin V2 in the 3′ UTR (figure 15-12) (78). It is conceivable that maturation protein binding to this site, which in Qβ is an HF binding site, inhibits replication and could thus serve as a regulatory signal. The MS2 RNA-maturation protein complex can be made in vitro and it is infectious for F^+ cells (78).

The second step in virion construction is the assembly of the coat protein around the RNA to form an icosahedral shell. This reaction does not depend on the A-protein. The nucleation point for this "coating" reaction is the replicase operator hairpin (figure 15-3). Coat protein will form capsids even in the absence of phage RNA. However, capsid formation occurs at a much lower concentration of coat protein if phage RNA is present, assuring that empty capsids will not form and all phage RNA becomes packaged (102, 103).

The infection cycle ends when the accumulation of the lysis protein in the cytoplasmic membrane has, in an indirect manner, caused the collapse of the cell wall. This 75-amino acid long hydrophobic protein appears to short-circuit the cytoplasmic membrane by forming pores (27). Loss of membrane potential then leads in an unknown way to degradation of peptidoglycan. In electron micrographs one sees usually that only a very small section of the sacculus (the cell wall network) has dissolved, often at the equatorial growth zone. Bacterial growth is required for cell lysis and in this respect there is a parallel with the action of penicillin (97).

There are no distinct motifs in the lysis proteins of the group A phages except for a strong clustering of hydrophobic amino acids near the C-terminus. The activity of the MS2 lysis protein is enclosed in the C-terminal 35 or so amino acids. The first 40 amino acids are dispensable (8a).

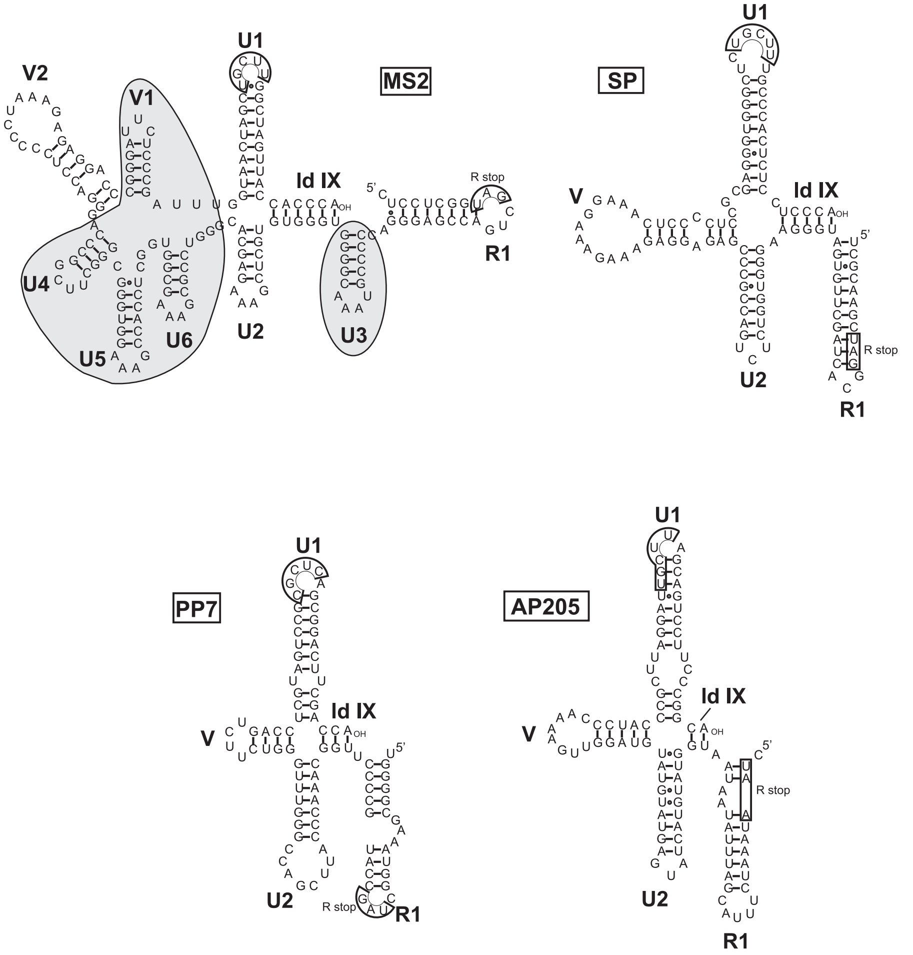


Figure 15-12 Comparison of RNA secondary structures in the 3′ UTR in four representative RNA phages. SP is a species IV allolevivirus. PP7 and AP205 resemble SP in their simple 3′ UTR structure. MS2 has a more elaborate folding with many extra elements (shaded). The stopcodon of replicase is boxed (R stop). Also boxed is a conserved sequence in the top of U1 (UGCUU). In Qβ and SP these nucleotides form a pseudoknot with their complement 1200 nucleotides upstream (figure 15-6). For MS2, PP7, and AP205 a similar interaction is supposed to exist.

There may be a similarity between the L-protein of the leviviruses and the gene E product of single-stranded DNA phage ΦX174 in the way these proteins induce lysis. The E-protein seems to interfere with the activity of the *E. coli* translocase I (MraY). This integral membrane protein catalyzes the formation of a murein precursor (11, 105).

Single-Stranded RNA Phages in Other Bacteria and Their Phylogenetic Relationships

For historical reasons the RNA phages of *E. coli* have received most attention and almost all of what we know has been derived from either Qβ or MS2 (R17, f2). Nevertheless, we may assume that very many, if not all, Gram-negative bacteria have their own RNA phages. Studying these will help us to determine which phage properties are general and which are specific for coliphages. In addition, such a comparative study may help to construct a plausible phylogenetic tree.

In the past single-stranded RNA phages have been isolated from *Caulobacter* and *Pseudomonas* (77). These infected their host via polar pili, making it clear that the F-pili pathway found for the coliphages is not a general feature. Lately, an RNA phage, AP205, growing in *Acinetobacter*, was discovered. The pilus used for infection has not been identified.

Only two non-coliphages have been sequenced: AP205 (43), and PP7 (59) which infects *Pseudomonas*. Nucleotide sequence identity among phages growing in different bacteria is close to random except for regions where conserved protein sequences are encoded, notably in parts of the replicase.

PP7 has the length and genetic map of MS2 or GA and was thus classified as a *Levivirus* (figure 15-1). There is no criterion to further assign PP7 to species I or II. AP205 on

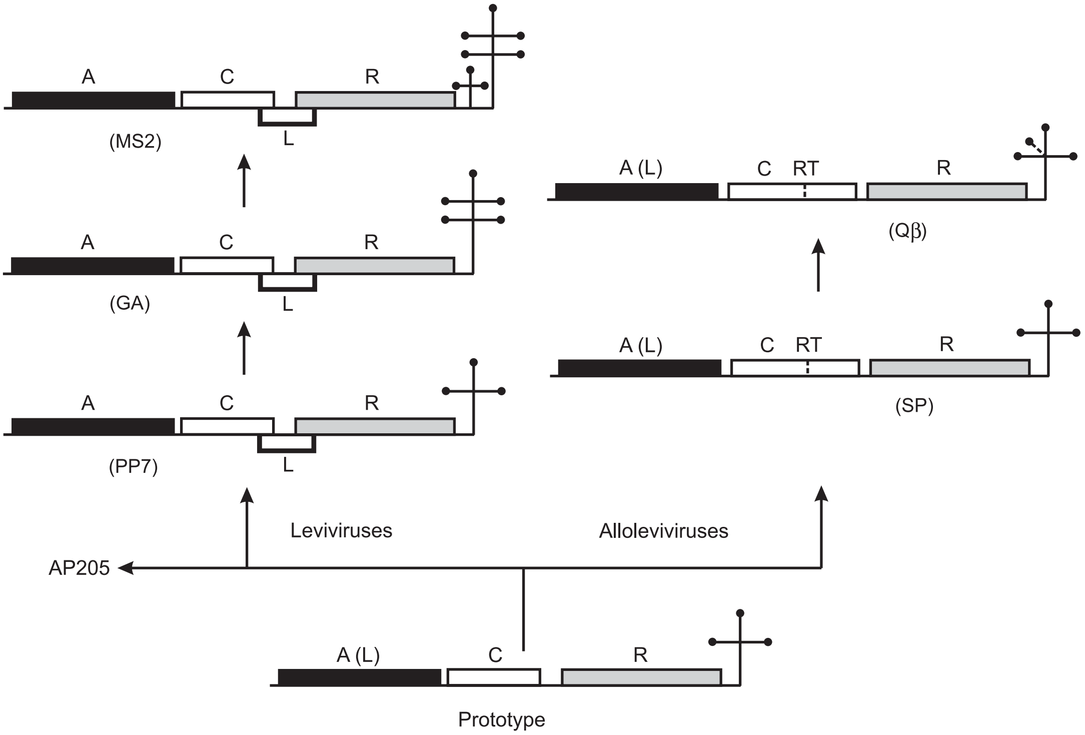


Figure 15-13 Proposed phylogenetic tree for the single-stranded RNA phages. The prototype combines infection and cell lysis in a single protein, A(L), and has a simple structure at the 3′ end as SP and PP7. The branching off of the alloleviviruses partly relieves protein A(L) from its dual role by creating the read-through protein which helps in infection. In the leviviruses the A(L) protein is relieved from its lysis function by evolving the dedicated lysis protein. In coliphages GA and MS2 we observe the development of increasingly elaborate 3′-end structures. A similar tendency can be seen between SP or NL95 (species IV) and Qβ, MX1, and M11 (species III).

the other hand is about as long as Qβ or SP, but considering that AP205 does not contain a read-through protein the phage was classified as a *Levivirus*. It is also unusual in the sense that its putative lysis gene occurred not at the classical position but appears at the 5′ terminus (ORF1 in figure 15-1).

Despite large differences in hosts and RNA sequences, common structural features can be recognized in all single-stranded RNA phages: (i) the extremely strong 5′ hairpin, (ii) relatively high single strandedness at the RBS of the coat gene, (iii) replicase operator (43), (iv) long-distance pseudoknot, (v) the folding of the 3′ UTR (see below) and (vi) a conserved pentanucleotide sequence in the loop of the 3′ terminal hairpin (figure 15-12) (43). There might even be a more extensive common pattern in the overall RNA folding, but at present we have not enough certainty about the RNA secondary structure of PP7 and AP205.

A most difficult question is the phylogenetic relationship between the known phages, especially between coliphages and non-coliphages and between leviviruses and alloleviviruses. Usually relationships are based on sequence identity between corresponding genes or proteins and this procedure is reliable when closely related phages need to be ordered or classified, for example to show that JP34 is species II and not I or that MX1 is species III rather than IV. However, when we must order MS2, Qβ, PP7, and AP205 the nucleotide or amino acid sequences provide little direction. The mutation rate of RNA phages is extremely high and as we have seen there can be many nucleotide changes within a few hours. For such and other reasons the RNA sequence may not reliably reflect the evolutionary history but rather show the "needs of the day," which may be quite different from those of yesterday or tomorrow. Therefore, to compare distantly related phages it may be better to look for features that are more stably inherited than nucleotide or amino acid sequence. One such stable feature is the genetic map. Another, as we have seen above, is the RNA folding. A conspicuous landmark in this respect is the structure of the 3′ UTR which, in the coliphages, is a hallmark of its genus and even species (2). Qβ and SP have a short simple version forming a 5- and 4-way junction, respectively. MS2 on the other hand has a much longer and complexer 3′ UTR (figure 15-12). Surprisingly, PP7 and AP205, though showing a *Levivirus* map, have the simple *Allolevivirus*-type 3′ end. Apparently, AP205 and PP7 take an intermediate position between MS2 and Qβ type phages.

Respecting genetic map and structure elements as deeply inherited traits and also assuming that evolution proceeds from simple to more complex, we arrive at the tree shown in figure 15-13. The proposed common ancestor has features that are common to all RNA phages known today. This prototype contains only the three major genes. There is neither read-through nor a separate lysis gene, and the 3′ UTR has the short simple version. In this prototype the lysis function can be carried out by the maturation protein as it still is today in Qβ. To arrive at alloleviviruses we need only a duplication or insertion (and its subsequent adaptation) to produce the extra read-through (RT) protein. On the other hand, to arrive at the present-day leviviruses it suffices to exploit a basic property of the bacterial ribosome, namely its restart capacity. By adjusting wobble positions in the early replicase gene, such restarts at the end of the coat gene can produce hydrophobic peptides, an important condition for lytic activity (MS2, GA, PP7). In AP205, lysis has evolved in a vacant part of the genome. As suggested (16), the pressure for these events may have come from the possibly inefficient double function of the maturation protein in the ancestor. The evolution of lysis and read-through protein can thus be considered as two different solutions for the same problem. In leviviruses the inefficient lysis property of the maturation protein is solved by developing a dedicated lysis gene, whereas in alloleviviruses the inefficient infection is improved by developing the read-through protein allowing maturation to become a better lysis protein. The simplicity of this scheme is attractive, even if it places the two coliphages MS2 and Qβ at the greatest possible evolutionary distance. The evolution of the elaborate 3′ UTR in MS2 and GA remains unexplained.

Acknowledgments

We would like to thank Dr. H. Weber for critical reading of this chapter and for permission to quote his unpublished results. Sjoerd van den Worm, René Olsthoorn, and Maarten de Smit kindly helped in preparing the figures.

References

1. Adhin, M. R., and J. van Duin. 1990. Scanning model for translational reinitiation in eubacteria. J. Mol. Biol. 213:811–818.
2. Adhin, M. R., J. Alblas, and J. van Duin. 1990. Secondary structure at the 3' terminal region of RNA coliphages: comparison with tRNA. Biochim. Biophys. Acta 1050:110–118.
3. Arora, R., C. Priano, A. B. Jacobson, and D. R. Mills. 1996. *Cis*-acting elements within an RNA coliphage genome: fold as you please, but fold you must!! J. Mol. Biol. 258:433–436.
4. Avota, E., V. Berzins, E. Grens, Y. Vishnevsky, R. Luce, and C. K. Biebricher. 1998. The natural 6S RNA (RQ RNA) found in Qβ-infected cells is derived from host and phage RNA. J. Mol. Biol. 276:7–17.
5. Axelrod, V. D., E. Brown, C. Priano, and D. R. Mills. 1991. Coliphage Qβ replication; RNA catalytic for single-stranded release. Virology 184:595–608.
6. Beekwilder, J. 1996. Secondary structure of the Qβ RNA. PhD. thesis. Leiden University, The Netherlands. Leiden, (Available on request.)
7. Beekwilder, J., R. Nieuwenhuizen, R. A. Poot, and J. van Duin. 1996. Secondary structure model for the first three domains of Qβ RNA. Control of A-protein synthesis. J. Mol. Biol. 256:8–19.
8. Beekwilder, J., and J. van Duin. Unpublished results.

8a. Berkhout, B., M. H. de Smit, R. A. Spanjaard, T. Blom, and J. van Duin. 1985. The amino terminal half of the MS-2 coded lysis protein is dispensable for function: implications for our understanding of coding region overlaps. EMBO J. 4:3315–3320.

9. Berkhout, B., and J. van Duin. 1985. Mechanism of translational coupling between coat and replicase genes in phage MS2. Nucleic Acids Res. 13:6955–6967.
10. Berkhout, B., B. F. Schmidt, A. van Strien, J. van Westrenen, and J. van Duin, J. 1987. Lysis gene of MS2 is activated by translation termination at the overlapping coat gene. J. Mol. Biol. 195:517–524.
11. Bernhardt, T. G., W. D. Roof, and R. Young. 2000. Genetic evidence that the bacteriophage ΦX174 lysis protein inhibits cell wall synthesis. Proc. Natl. Acad. Sci. USA 97:4297–4302.
12. Biebricher, C. K., M. Eigen, and W. C. Gardiner. 1984. Kinetics of RNA replication: Plus–minus asymmetry and double strand formation. Biochemistry 23:3186–3194.
13. Biebricher, C. K., and R. Luce. 1993, Sequence analysis of RNA species synthesized by Qβ replicase without template. Biochemistry 32:4848–4854.
14. Biebricher, C. K. 1999. Mutation, competition and selection as measured with small RNA molecules, p. 65. *In* E. Domingo (ed.) Origin and Evolution of Viruses. Academic Press, London.
15. Blumenthal, T., and G. G. Carmichael. 1979. RNA replication: function and structure of Qβ replicase. Annu. Rev. Biochem. 48:525–548.
16. Bollback, J. P., and J. P. Huelsenbeck. 2001. Phylogeny, genome evolution and host specificity of ss RNA bacteriophage (Family *Leviviridae*). J. Mol. Evol. 52:117–128.
17. Boni, I. V., D. M. Isaeva, and E. L. Budowski. 1986. Ribosomal protein S1 binds to an internal region of the replicase gene within the complex of a 30S ribosomal subunit and MS2 RNA. Bioorg. Khim. 12:293 (in Russian).

17a. Boni, I. V., D. M. Isaeva, and M. L. Musychenko. 1991. Ribosome–messenger recognition: mRNA target sites for ribosomal protein S1. Nucleic Acids Res. 19:155–162.

18. Brown, D., and L. Gold. 1995. Template recognition by an RNA-dependent RNA polymerase: identification and characterization of two RNA-binding sites on Qβ replicase. Biochemistry 34:14765–14774.
19. Brown, D., and L. Gold. 1995. Selection and characterization of RNAs replicated by Qβ replicase. Biochemistry 34:14775–14782.

20. Brown, D., and L. Gold. 1996. RNA replication by Qβ replicase: a working model. Proc. Natl. Acad. Sci. USA 93:11558–11562.
21. Brown, L., and T. Elliott. 1997. Mutations that increase expression of the *rpoS* gene and decrease its dependence on *hfq* function in *Salmonella typhimurium*. J. Bacteriol. 17:656–662.
22. Chetverin, A. B., and A. S. Spirin. 1995. Replicable RNA vectors: prospects for cell-free gene amplification, expression and cloning. Progr. Nucleic Acids Res. Mol. Biol. 51:225–270.
23. Court, D. 1993. RNA processing and degradation by RNase III, p. 71. *In* J. G. Belasco and Brouwerman (eds.) Control of RNA Stability. Academic Press, New York.
24. de Smit, M. H., and J. van Duin. 1990. Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. Proc. Natl. Acad. Sci. USA 87:7668–7672.
25. de Smit, M. H. 1998. Translational control by mRNA structure in eubacteria: molecular biology and physical chemistry. *In* M. Grunberg-Manago and R. Simons, (eds.) RNA Structure and Function. Cold Spring Harbor Laboratory press, Cold Spring Harbor, N.Y.
25a. de Smit, M. H., and J. van Duin. 2003. Translational standby sites: how ribosomes may deal with the rapid folding kinetics of mRNA. J. Mol. Biol. 331:737–743.
26. Furuse, K. 1987. Distribution of coliphages in the environment: general considerations, p. 87. *In* S. M. Goyal, (ed.), Phage Ecology, Wiley, New York.
27. Goessens, W. H. F., A. J. M. Driessen, J. Wilschut, and J. van Duin. 1988. A synthetic peptide corresponding to the C-terminal 25 residues of phage MS2 lysis protein dissipates the proton motive force in *E. coli* membrane vesicles by generating hydrophilic pores. EMBO J. 7:867–873.
28. Golmohammadi, R. K. Fridborg, M. Bundule, K. Valegård, and J. Liljas. 1996. The crystal structure of phage Qβ at 3.5 Å resolution. Structure 4:543–554.
29. Grahn, E., N. J. Stonehouse, J. B. Murray, S. van den Worm, K. Valegård, K. Fridborg, P. G. Stockley, and L. Liljas. 1999. Crystallographic studies of RNA hairpins in complexes with recombinant MS2 capsids: implications for binding requirements. RNA 5:131–138.
30. Groeneveld, H., K. Thimon, and J. van Duin. 1995. Translational control of maturation protein synthesis in phage MS2: a role for the kinetics of RNA folding. RNA 1:79–88.
31. Jacobson, A. B., and M. Zuker. 1993. Structural analysis by energy dot plot of a large mRNA. J. Mol. Biol. 233:261–269.
32. Inokuchi, Y., R. Takahashi, T. Hirose, S. Inayama, A. B. Jacobson, and A. Hirashima. 1986. The complete nucleotide sequence of group II phage GA. J. Biochem. 99:1169–1180.
33. Inokuchi, Y., A. B. Jacobson, T. Hirose, S. Inayama, and A. Hirashima. 1988. Analysis of the complete nucleotide sequence of group IV coliphage SP. Nucleic Acids Res. 16:6205–6221.
34. Inokuchi, Y., and M. Kajitani. 1997. Deletion analysis of Qβ replicase. J. Biol. Chem. 272:15339–15345.
35. Jacobson, A. B., R. Arora, M. Zuker, C. Priano, C. H. Liu, and D. R. Mills. 1998. Structural plasticity in RNA and its role in the regulation of translation in Qβ. J. Mol. Biol. 275:589–600.
36. Kajitani, M., A. Kato, A. Wada, Y. Inokuchi, and A. Ashihama. 1994. Regulation of the *E. coli hfq* gene encoding the host factor for phage Qβ. J. Bacteriol. 176:531–534.
37. Karnik, S., and M. Billeter. 1983. The lysis function of RNA bacteriophage Qβ is mediated by the maturation protein. EMBO J. 2:1521–1526.
37a. Karring, H., S. Mathu, J. Van Duin, B. Clark, B. Kraal, and C. Knudsen. 2004. Qβ-phage resistance by deletion of the coiled-coil motif in elongation factor Ts. J. Biol. Chem. 279:1878–1884.
38. Kawashima, T., G. Berthet-Colominas, H. Wulff, S. Cusack, and R. Leberman. 1996. The structure of the *E. coli* EF-Tu·EF-Ts complex at 2.5 Å resolution. Nature 379:511–518.
39. Klovins, J., J. van Duin, and R. C. L. Olsthoorn. 1997. Rescue of the RNA phage genome from RNase III cleavage. Nucleic Acids Res. 25:4201–4208.
40. Klovins, J., N. V. Tsareva, M. H. de Smit, V. Berzins, and J. van Duin. 1997. Rapid evolution of translational control mechanisms in RNA genomes. J. Mol. Biol. 265:372–384.
41. Klovins, J., V. Berzins, and J. van Duin. 1998. A long-range interaction that bridges the thousand nucleotides between the S-site and the 3′ end is required for replication. RNA 4:948–957.
42. Klovins, J., and J. van Duin. 1999. A long-range pseudoknot in Qβ RNA is essential for replication. J. Mol. Biol. 294:875–884.
43. Klovins, J., G. P. Overbeek, S. van den Worm, H.-W. Ackermann, and J. van Duin. 2002. Nucleotide sequence of a ss RNA phage from *Acinetobacter*: kinship to coliphages. J. Gen. Virol. 83:1523–1533.
44. Knolle, P., and T. Hohn. 1975. Morphogenesis of RNA phages, p. 147. *In* N. Zinder (ed.) RNA Phages. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
45. Kolakofsky, D., M. A. Billeter, K. Weber, and C. Weissmann. 1973. Resynchroniation of RNA synthesis by Qβ replicase at the internal site of the template. J. Mol. Biol. 76:271–284.
45a. Kondo, M. 1975. Structure and function of Qβ replicase. Arch. Int. Physiol. Biochem. 83:909–948.
46. Licis, N., Balklava, Z., and J. van Duin. 2000. Forced retroevolution of an RNA bacteriophage. Virology 271:298–306.
47. Lindner, A. J., S. J. Glaser, C. K. Biebricher, and G. R. Hartman. 1991. Self-catalyzed affinity labeling of Qβ replicase. Eur. J. Biochem. 202:249–254.
48. Mathu, S. G. J., C. R. Knudsen, J. van Duin, and B. Kraal. 2003. Isolation of Qβ polymerase complexes containing mutant species of EF-Tu. J. Chromatogr. 786:279–286.
49. Mathu, S. G. J., H. Karring, C. R. Knudsen, J. van Duin, and B. Kraal, manuscript in preparation.
50. Majdolani, N., C. Cunning, D. Sledjeski, T. Elliott, and S. Gottesman. 1998. DsrA RNA regulates translation of rpoS message by an anti-antisense mechanism,

independent of its action as an antisilencer of transcription. Proc. Natl. Acad. Sci. USA 95:12462–12467.

51. Meyer, F., H. Weber, and C. Weissmann. 1981. Interactions of Qβ replicase with Qβ RNA. J. Mol. Biol. 153:631–660.
52. Miranda, G., D. Schuppli, L. Barrera, C. Hausherr, J. M. Sogo, and H. Weber. 1997. Recognition of Qβ (+) strand as a template by Qβ replicase: role of interactions mediated by S1 and HF. J. Mol. Biol. 267:1089–1103.
53. Moody, M., J. L. Burg, R. DiFrancesc, D. Lovern, W. Stanick, J. Lin-Goerke, K. Mahdavi, Y. Wu, and M. P. Farrell. 1994. Evolution of host cell RNA into efficient template RNA by Qβ replicase: the origin of RNA in untemplated reactions. Biochemistry 33:13836–13847.
54. Munishkin, A. V., L. A. Voronin, and A. B. Chetverin, A. B. 1988. An in vivo recombinant RNA capable of autocatalytic synthesis by Qβ replicase. Nature 333:473–475.
55. Munishkin, A. V., L. A. Voronin, V. L. Ugarov, L. A. Bandareva, H. V. Chetverina, and A. B. Chetverin. 1991. Efficient templates for Qβ replicase are formed by recombination from heterologous sequences. J. Mol. Biol. 221:463–472.
56. Murphy, F. A., C. M. Fouquet, D. H. L. Bishop, S. A. Ghabrial, A. W. Jarvis, M. A. Mayo, and M. D. Summers. 1995. Virus Taxonomy: The Classification and Nomenclature of Viruses. Springer, Berlin.
57. Nissen, P., S. Thirup, M. Kjeldgaard, and J. Nyborg. 1999. The crystal structure of Cys-tRNA^{cys}-EF-Tu-GDPNP reveals general and specific features in the ternary complex and in tRNA. Structure Fold Des. 7:143–156.
58. Olsthoorn, R. C. L., N. Licis, and J. van Duin. 1994. Leeway and constraints in the forced evolution of a regulatory RNA helix. EMBO J. 13:2660–2668.
59. Olsthoorn, R. C. L., G. Garde, T. Dayhuff, J. F. Atkins, and J. van Duin. 1995. Nucleotide sequence of a ss RNA phage from *Pseudomonas aeruginosa*: kinship to coliphages and conservation of regulatory structures. Virology 206:611–625.
60. Olsthoorn, R. C. L., and J. van Duin. 1996. Random removal of inserts from an RNA genome: selection against single-stranded RNA. J. Virol.. 70:729–736.
61. Olsthoorn, R. C. L., and J. van Duin. 1996. Evolutionary reconstruction of a hairpin deleted fom the genome of an RNA virus. Proc. Natl. Acad. Sci. USA 93:12256–12261.
62. Palasingam, K., and P. N. Shaklee. 1992. Reversion of Qβ RNA phage mutants by homologous RNA recombination. J. Virol.. 66:2435–2442.
63. Paranchych, W. 1975. Attachment, ejection and penetration stages of the RNA phage infection process, p. 85. *In* N. Zinder (ed.) RNA Phages. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
64. Paranchych, W., and L. S. Frost. 1988. The physiology and biochemistry of pili. Adv. Microbial Physiol. 29:53–114.
65. Pavlov, M. Y., D. V. Freistroffer, J. McDougal, R. H. Buckingham, and M. Ehrenberg. 1997. Fast recycling of *E. coli* ribosomes requires both RRF and release factor RF3. EMBO J. 16:4134–4141.
66. Peabody, D. S. 1997. Role of the coat protein–RNA interaction in the life cycle of bacteriophage MS2. Mol. Gen. Genet. 254:358–364.
67. Poot, R. A., N. V. Tsareva, I. V. Boni, and J. van Duin. 1997. RNA folding kinetics regulates translation of phage MS2 maturation gene. Proc. Natl. Acad. Sci. USA 94:10110–10115.
68. Preuss, R., J. Dapprich, and N. G. Walter. 1997. Probing RNA–protein interactions using pyrene-labeled oligodeoxynucleotides: Qβ replicase efficiently binds small RNAs by recognizing pyrimidine residues. J. Mol. Biol. 273:600–613.
69. Schmidt, B. F., B. Berkhout, G. Overbeek, A. van Strien, and J. van Duin. 1987. Determination of the RNA structure that regulates lysis gene expression in MS2. J. Mol. Biol. 195:505–516.
70. Schuppli, D., L. Barrera, and H. Weber. 1994. Identification of recognition elements on bacteriophage Qβ minus strand RNA that are essential for template activity with Qβ replicase. J. Mol. Biol. 243:811–815.
71. Schuppli, D., G. Miranda, H.-C. T. Tsui, M. E. Winkler, J. H. Sogo, and H. Weber. 1997. Altered 3′ terminal structure in phage Qβ adapted to host factor-less *E. coli*. Proc. Natl. Acad. Sci. USA 94:10239–10242.
72. Schuppli, D., G. Miranda, S. Qiu, and H. Weber. 1998. A branched stem-loop structure in the M-site of Qβ RNA is important for template recognition by Qβ replicase holoenzyme. J. Mol. Biol. 283:585–593.
73. Schuppli, D., J. Georgijevic, and H. Weber. 2000. Synergism of mutations in phage Qβ RNA affecting host factor dependence of Qβ replicase. J. Mol. Biol. 295:149–154.
74. Senear, A. W., and J. A. Steitz. 1976. Site specific interaction of Qβ host factor and ribosomal protein S1 with Qβ and R17 bacteriophage RNA. J. Biol. Chem. 251:1902–1912.
75. Shaklee, P. N., J. J. Miglietta, A. C. Palmenbrg, and P. Kaesberg. 1988. Infectious positive and negative-strand transcript RNAs from Qβ cDNA clones. Virology 163:209–213.
76. Shaklee, P. N. 1990. Negative-strand RNA replication by Qβ and MS2 positive-strand RNA phage replicases. Virology 178:340–343.
77. Shapiro, L., and L. Bendis. 1975. RNA phages of bacteria other than *E. coli*, p. 397. *In* N. Zinder (ed.) RNA Phages, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
78. Shiba, T., and Y. Suzuki. 1981. Localization of A protein in the RNA–A protein complex of RNA phage MS2. Biochim. Biophys. Acta. 654:249–255.
79. Skripkin, E. A., and A. B. Jacobson. 1993. A 2D model at the nucleotide level for the central hairpin in Qβ RNA. J. Mol. Biol. 233:245–260.
80. Sledjeski, D., C. Whitman, and A. Zhang. 2001. Hfq is necessary for regulation by the untranslated RNA DsrA. J. Bacteriol. 183:1997–2005.
81. Sørensen, M. A., J. Fricke, and S. Pedersen, S. 1998. Ribosomal protein S1 is required for translation of most, if not all, natural mRNAs in *E. coli*. J. Mol. Biol. 280:561–569.
82. Steege, D. A. 2000. Emerging features of mRNA decay in bacteria. RNA 6:1079–1090.
83. Taniguchi, T., M. Palmieri, and C. Weissmann. 1978. Qβ DNA-containing plasmids giving rise to Qβ phage formation in the bacterial host. Nature 274:223–228.

84. Tars, K., M. Bundule, K. Fridborg, and L. Liljas. 1997. The crystal structure of phage GA and a comparison of bacteriophages belonging to the major groups of *E. coli* leviviruses. J. Mol. Biol. 271:759–773.
85. Tars, K., Fridborg, K., Bundule, M., and Liljas, L. 2000. The three-dimensional structure of phage PP7 from *Pseudomonas aeruginosa* at 3.7 Å resolution. *Virology* 272:331–337.
86. Tretheway, D. M., S. Joshinari, and T. W. Dreher. 2001. Autonomous role of 3′ terminal CCCA in directing transcription of RNAs by Qβ replicase. J. Virol. 75:11373.
87. Tsareva, N. V., and J. van Duin. Unpublished results.
88. Valegård, K., L. Liljas, K. Fridborg, and T. Unge. 1990. The three-dimensional structure of the bacterial virus MS2. Nature 345:36–41.
89. Valegård, K., J. B. Murray, P. G. Stockley, N. J. Stonehouse, and L. Liljas. 1994. Crystal structure of an RNA phage coat protein–operator complex. Nature 371: 623–626.
90. Valegård, K., J. B. Murray, N. J. Stonehouse, S. van den Worm, P. G. Stockley, and L. Liljas. 1997. The 3D structures of two complexes between recombinant MS2 capsids and RNA operator fragments. J. Mol. Biol. 270:724–738.
91. van Dieijen, G., P. Zipori, W. van Prooijen, and J. van Duin. 1978. Involvement of ribosomal protein S1 in the assembly of the initiation complex. Eur. J. Biochem. 90:571–580.
92. van Duin, J. 1988. The Single-Stranded RNA Bacteriophages, p. 117. *In* R. Calendar (ed.) The Bacteriophages. Plenum Press, New York.
93. van Himbergen, J., B. van Geffen, and J. van Duin. 1993. Translational control by a long-range RNA–RNA interaction; a basepair substitution analysis. Nucleic Acids Res. 21:1713–1717.
94. van Meerten, D., M. Zelwer, P. Régnier, and J. van Duin. 1999. In vivo oligo(A) insertions in phage MS2: role of *E. coli* poly(A) polymerase. Nucleic Acids Res. 27:3891–3898.
95. van Meerten, D., G. Girard, and J. van Duin. 2001. Translational control by delayed RNA folding; identification of the kinetic trap. RNA 7:483–494.
96. van Meerten, D., and J. van Duin. 2002. Levivirus. *In* Springer Index of Viruses. Springer, Berlin.
97. Walderich, G., A. Ursinus-Wossner, J. van Duin, and J.-V. Höltje. 1988. Induction of the autolytic system of *E. coli* by specific insertion of phage MS2 lysis protein into the bacterial cell envelope. J. Bacteriol. 170:5027–5033.
98. Weber, H., M. A. Billeter, S. Kahane, C. Weissmann, J. Hindley, and A. Porter. 1972. Molecular basis for repressor activity of Qβ replicase. Nature 237:166–170.
99. Weber, H. 1999. Qβ replicase, p. 2085. *In* T.E. Creighton (ed.) The Encyclopedia of Molecular Biology, Wiley, New York.
100. White, A.K., and J. T. Morris. 1995. RNA determinants of junction site selection in RNA virus recombinants and defective interfering RNAs. RNA 1:1029–1040.
101. Winter, R. B., and L. Gold. 1983. Overproduction of Qβ maturation protein leads to cell lysis. Cell 33:877–885.
102. Witherell, G. W., H. Wu, and O. C. Uhlenbeck. 1990. Cooperative binding of R17 coat protein to RNA. Biochemistry 29:11051–11057.
103. Witherell, G. W., J. M. Gott, and O. C. Uhlenbeck. 1991. Specific interactions between RNA phage coat protein and RNA. Prog. Nucleic Acids Res. Mol. Biol. 40:185–220.
104. Yonesaki, T., K. Furuse, L. Haruna, and I. Watanabe. 1982. Relationships among four groups of RNA coliphages based on template specificity of GA replicase. Virology 116:379–381.
105. Young, R., I.-N. Wang, and W. D. Roof. 2002. Phages will out: strategies of host cell lysis. Trends Microbiol. 8:120–128.
106. Zamora, H., R. Luce, and C. K. Biebricher. 1995. Design of artificial short-chained RNA species that are replicated by Qβ replicase. Biochemistry 34:1261–1266.
107. Zinder, N. D. 1975. Preface *In* RNA Phages, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

16

Phages with Segmented Double-Stranded RNA Genomes

LEONARD MINDICH

The first member of this family of bacteriophages, the *Cystoviridae*, was isolated in 1973. This phage, named ϕ6, was found to have a genome of three double-stranded RNA molecules packaged within a polyhedral nucleocapsid surrounded by a lipid-containing membrane (47, 52). The basic outlines of the life cycle of ϕ6 were described in several reviews (25, 29). Since that time, a number of important advances have been made. The structures of the virus and subviral particles have been defined through the application of cryo-electron microscopy (1, 7). The structure of the RNA-dependent polymerase has been elucidated through X-ray diffraction studies on protein crystals (2, 3). The reassembly of inner core particles of both ϕ6 and ϕ8 has been accomplished with pure individual proteins (21, 40). The packaging of the genomic segments has been elucidated and several reverse genetic systems have been developed (27, 28, 48). Recombination between the genomic segments has been described (26). In addition, a number of relatives of ϕ6 have been isolated from nature, enlarging the family of *Cystoviridae* (31).

The Family of *Cystoviridae*

Bacteriophage ϕ6 and other members of the *Cystoviridae* have genomes of three double-stranded RNA segments. They are designated L, M, and S and their respective sizes are about 6.5 kbp, 4 kbp, and 3 kbp (figure 16-1). The L segment contains the genes for the proteins of the inner core structure, which is also called the polymerase complex or the procapsid. Segment M contains genes for host attachment and segment S contains genes for membrane structure and assembly as well as lytic proteins and protein P8, which in some cases forms a shell around the inner core. The inner core is a dodecahedron when empty and is composed of four proteins: P1, P2, P4, and P7 (figure 16-2). P1 is the major structural protein of the inner core and 120 molecules of P1 are arrayed in a pattern that is common to most dsRNA viruses that infect eukaryotic cells. P2 is the RNA dependent RNA polymerase.

All the members of the *Cystoviridae* except ϕ8 have a shell of P8 around the filled inner core. In the case of ϕ8, protein P8 is a component of the membrane. All members of the family have a membrane as the exterior structure of the virion. ϕ6 and its closest relatives have a host attachment structure composed of a hydrophobic protein P6 that anchors the attachment specificity protein P3 to the membrane. These phages attach to a specific type IV pilus on the host bacterium. The other members of the family—ϕ8, ϕ12 and ϕ13—attach to rough lipopolysaccharide (LPS) on the host cells (31). This is mediated by a structure composed of a number of large proteins designated as P3a, P3b, and/or P3c. There is no sequence similarity between the two classes of proteins. Although the host specificity of the phages is primarily to pseudomonads, particularly *Pseudomonas syringae*, mutants of the phages can plaque on other pseudomonads such as *Pseudomonas pseudoalcaligenes*. The phages that attach to rough LPS can infect rough LPS mutants of other Gram-negative bacteria, but do not form plaques on them. However, ϕ8 can form plaques at high efficiency on heptoseless mutants of *Salmonella typhimurium* (17). The family *Cystoviridae* now consists of nine isolates. It seems reasonable that there are many more members in nature and that the host range will be extended to other Gram-negative bacteria. Although these phages are somewhat more fragile than the double-stranded DNA phages, they appear to have found a niche that becomes available after a population has responded to infection by phages that attach to normal LPS. ϕ6 was isolated from bean plants; a number of the newer phages were isolated from snow pea leaves and others were isolated from the leaves of carrot, radish, basil, and tomato plants (31).

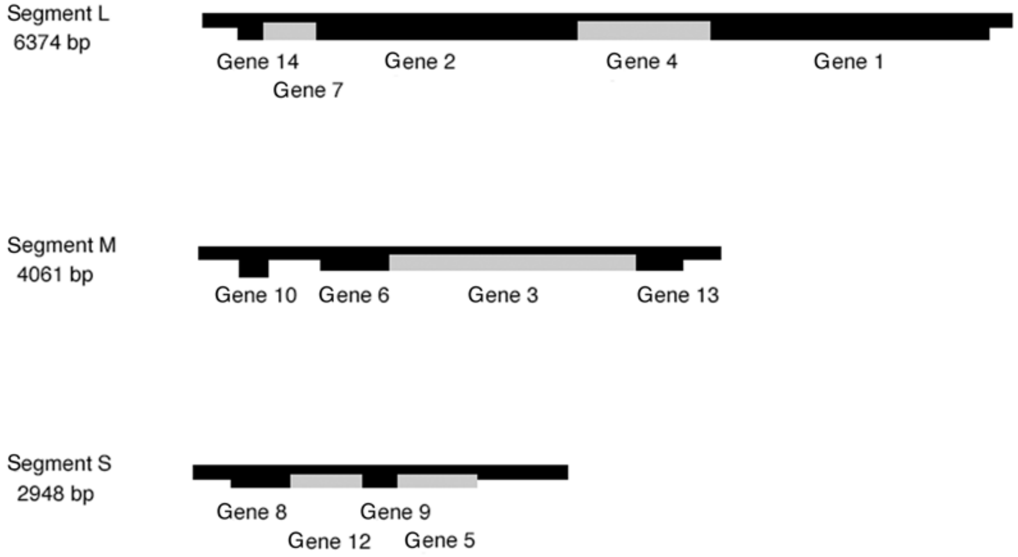


Figure 16-1 The genomic segments of ϕ6. Proteins P7, P2, P4, and P1 comprise the inner core particle. Proteins P6 and P3 are the host attachment apparatus while P10 and P5 are involved in host cell lysis. P5 is also involved in host penetration. P9 is the major membrane protein, which is inserted into membrane through the action of P12. P8 forms a shell around the inner core particle. Each segment has a unique *pac* region near the 5′ end of the (+) strand.

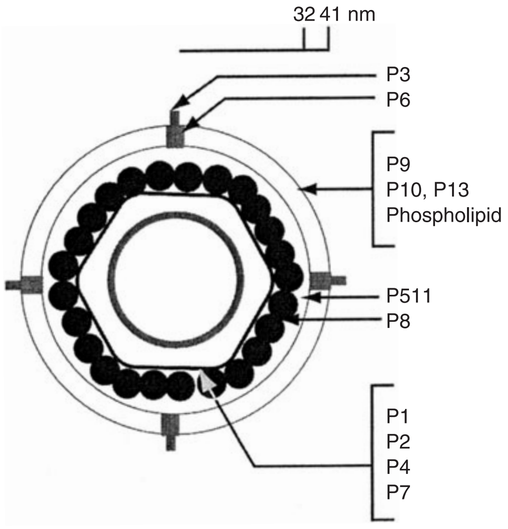


Figure 16-2 A simplified diagram of cystovirus structure. double-stranded RNA is contained inside a core of P1, P2, P4, and P7. This structure is within a shell of P8 (51) in all members except ϕ8, where P8 is a membrane component. The lipid-containing membrane surrounds the P8 shell. Attachment proteins P6 and P3 are in the membrane along with P9 (the major membrane protein) and P10. See thebacteriophages.org/frames_0160.htm for a color version of this figure.

Structure of the Inner Core

Four proteins coded by segment L assemble to form the procapsid or inner core of these viruses. The proteins are designated P1, P2, P4, and P7. P1 is the major structural protein and is present in 120 copies (6, 7). P2 is the RNA-dependent RNA polymerase, responsible for both (+) and (−) strand synthesis (14, 53). There are probably 12 copies per virion. P4 is an NTPase that is necessary for the packaging of (+) strand RNA molecules into the procapsid (13). It appears as a homohexamer at each of the 12 5-fold faces of the ϕ6 procapsid (7), but in lower proportions in the ϕ8 procapsid (49). Protein P7 plays an auxiliary role in assembly, packaging, and RNA synthesis (14, 19). The empty procapsid has a polyhedral shape with 5-fold vertices pulled toward the center of the structure (7). Filled particles have a more spherical shape, with the P4 turrets extended (figure 16-3) (1).

The procapsid structure is somewhat unstable and freeze-thawing in high salt leads to it falling apart into its constituent proteins. It has been possible to isolate the individual proteins and then to reassemble the particles at high efficiency (21, 40). In this manner it has been possible to study the kinetics of assembly of the core and to reconstitute infectious virions from the component proteins and RNA. The kinetics of assembly suggest that nucleation of ϕ6 and ϕ8 involves tetramers of protein P1 interacting with P4 hexamers and, in the case of ϕ8, molecules of P2.

Genomic Packaging

Purified procapsids can be isolated from cultures of *E. coli* carrying plasmids with cDNA copies of genomic segment L. These structures are capable of recognizing cognate RNA, packaging the RNA, replicating it as (−) strand to form double-stranded RNA and finally to form (+) strand

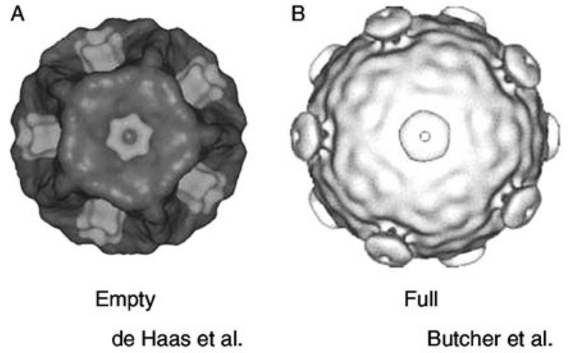


Figure 16-3 Phage ϕ6 inner core particles. A: Cryo-electron microscopy reconstructions of empty procapsids (7). B: Filled procapsids (1). The structures on the 5-fold faces are the multimers of the NTPase, P4.

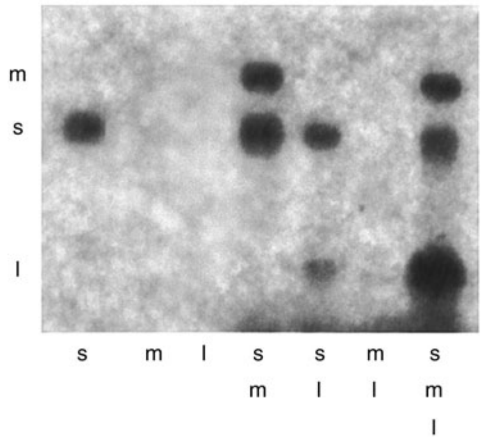


Figure 16-5 Serial dependence of genomic packaging. Packaging of radioactive (+) strands of exact copies of genomic segments S and M and of a truncated segment L of ϕ6. Radioactive transcripts of plasmids were incubated with procapsids, treated with RNase I, and applied to a 2% agarose gel.

messenger RNA which is extruded from the particle (27). All these reactions can take place in vitro with no accessory factors from the host cells. The particles can even carry out recombination between the RNA molecules (26).

Using the in vitro system, it has been possible to work out the rules governing genomic packaging. (+) strand RNA is packaged in the order s:m:l with a requirement for hydrolysis of NTP. The NTP efficacy in packaging reflects the activity for NTPase of the hexameric form of protein P4 in ϕ6 (13, 20). The sequences are identical for 18 bases in ϕ6 with the exception of the second nucleotide, but vary in the other phages (figure 16-4). The 5′ end of the (+) strands has a sequence that is very AU-rich. This is probably to facilitate melting for transcription. Each of the (+) strand copies has a *pac* sequence near, but not at the 5′ end. These sequences in ϕ6 are approximately 200 nucleotides in length. They are completely different for each segment with virtually no sequence identity between them (12). These sequences promote specific secondary structures with a number of stem-loops (27, 39). The (+) strands also have stem-loop structures at their 3′ ends. These structures are primarily for nuclease resistance but the terminal sequences do play a role in polymerase specificity (35). The (+) strand molecules of S can be packaged alone, but M requires the prior packaging of S and packaging of L requires the prior packaging of M (figure 16-5). (−) strand synthesis begins when the full complement of RNA is packaged (10). (+) strand synthesis begins when the entire genome is converted to double-stranded RNA. The (+) strands are packaged from the 5′ ends. RNA with strong hairpins at the 3′ end is packaged but the hairpin structures remain outside the particle (44). Studies with isolated hexamers of P4 of ϕ8 and ϕ12 show that RNA passes through the hexamer in a 5′ to 3′ direction (22).

ϕ6	l	GUAAAAAAACUUUAUAUA
	m, s	GGAAAAAAACUUUAUAUA
ϕ13	l	GAUAAAAAACUUUAUAUU
	m	GGAAAAAAACUUAUAUUU
	s	GGAAAAAACUUUUAUAA
ϕ12	l	ACAAUAAAAUAAUACAA
	m	GAAUUAAUUAAAUUUUA
	s	GAAUAAACUAAAUAAAG
ϕ8	l	GAAAUUUAUUUAAAGCU
	m	GAAAUUUUCAAAGUCUU
	s	GAAAUUUUCAAAUCUUU

Figure 16-4 5′ sequences of (+) strand transcripts of genomic segments.

A model that rationalizes the peculiarities of packaging has been proposed (figure 16-6) (27). It suggests a number of predictions that have been borne out experimentally. The model proposes that the outside of the empty procapsid has binding sites for the (+) strand of segment S. The RNA binds and its 5′ end is positioned in the pore of a P4 hexamer. The NTPase acts as a motor to pull the RNA into the particle. As a result of this uptake, the particle expands and the binding sites for S are lost and binding sites for M appear. The process is repeated, resulting in the loss of the binding sites for M and the appearance of binding sites for L. Once segment L is packaged, another conformational change results in the activation of the polymerase to begin (−) strand synthesis. Upon completion of (−) strand synthesis, the RNA content of the particle is doubled, resulting in a further conformational change in the particle, resulting in the onset of (+) strand transcription.

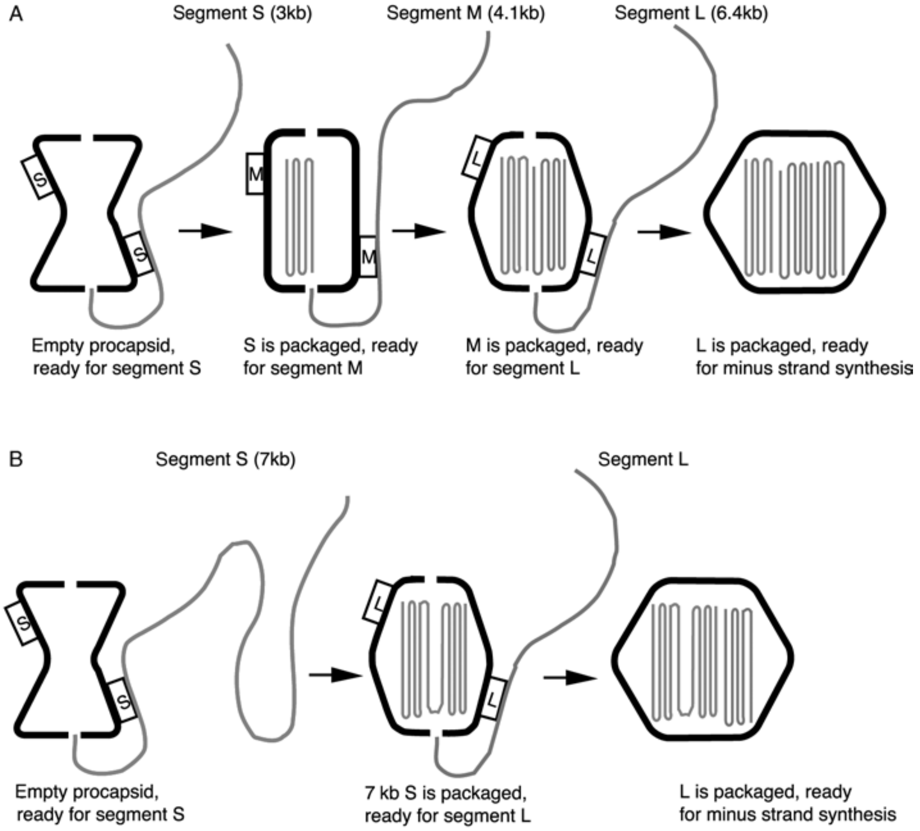

A
Segment S (3kb)
Segment M (4.1kb)
Segment L (6.4kb)
S
S
M
M
L
L
Empty procapsid, ready for segment S
S is packaged, ready for segment M
M is packaged, ready for segment L
L is packaged, ready for minus strand synthesis
B
Segment S (7kb)
Segment L
S
S
L
L
Empty procapsid, ready for segment S
7 kb S is packaged, ready for segment L
L is packaged, ready for minus strand synthesis

Figure 16-6 The packaging model. A: The empty procapsid shows only binding sites for S. After a full-size S is packaged, the S sites disappear and M sites appear. After a full-size M is packaged, the M sites disappear and L sites appear. After a full-size L is packaged, (−) strand synthesis commences. After (−) strand synthesis is completed, (+) strand synthesis commences. B: If segment S is of the size equal to the sum of both S and M, the S sites will disappear and the L sites will appear and segment L will be packaged without segment M. See thebacteriophages.org/frames_0160.htm for a color version of this figure.

Packaging of RNA can be studied easily in the ϕ6 system because RNA that is inside the core particle is resistant to RNase I (9, 41). Packaging of a normal (+) strand molecule can be inhibited by excess amounts of smaller homologous RNA as long as the competing RNA has a proper *pac* sequence (45). However, it is possible to construct competing RNA molecules that cannot themselves be packaged. Reducing the distance between the 18 base consensus sequence at the 5′ end and the *pac* sequence does this. If the deletion is too great, the RNA does not compete or package; but there is a sequence reduction where the RNA is not packaged but does compete. This supports the contention that the specificity of packaging is determined on the outside of the particle. Additional studies have shown that binding of RNA to the surface of ϕ6 procapsids is specific in the presence of ATP (42). The (+) strand of S binds well but those of M and L do not bind under normal conditions. Crosslinking of RNA to procapsids has shown an association with the region between amino acids 98 and 155 in P1, the major structural protein of the inner core (42). P1 of ϕ6 has 769 amino acids.

RNA molecules that are smaller than their normal size are packaged as long as they contain a proper *pac* sequence; however, the packaging system brings in an amount of RNA that approximates the normal weight of that RNA segment. If the molecule is half-size, then two are packaged; if it is quarter-size then four molecules are packaged. This is consistent with the idea that packaging specificity changes as a result of the expansion of the procapsid surface due to the amount of RNA packaged. It was predicted that a molecule with the *pac* sequence of S with the size of S plus M would promote the packaging of segment L without the requirement for segment M. This was borne out. Similarly, it was found that packaging a normal S segment along with a segment with the M *pac* sequence but the size of M plus L would turn on (−) strand synthesis without the requirement of RNA with L sequence. The procapsids were also capable of packaging an RNA molecule equal in size to the entire genome of ϕ6, 14 kb, which then promoted (−) strand synthesis followed by (+) strand synthesis.

These finding were corroborated with experiments in which living virus was constructed with genomic segments

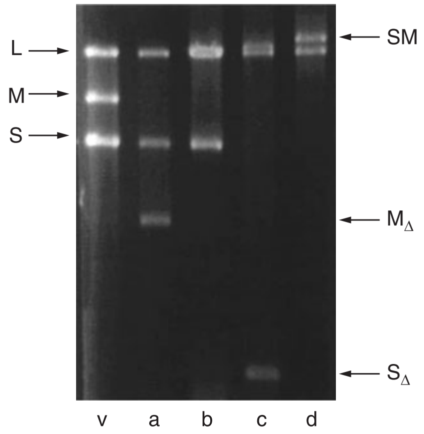


Figure 16-7 Agarose gel electrophoresis of double-stranded RNA isolated from virions. Lane V shows the distribution of normal segments L, M, and S. Lane a shows double-stranded RNA from bacteriophage ϕ2007, which has a deletion in segment M. Lane b shows RNA from ϕ2064, which has normal L, an MS chimera picked up from pLM1114, and a normal segment S. Lane c shows RNA from ϕ2323, which contains normal L, the MS chimera shown in b, and a deleted segment S that contains no genes and is only 798 bp. Lane d shows RNA from ϕ2361, which contains normal L, a chimera of S and M, but no normal segment M or S.

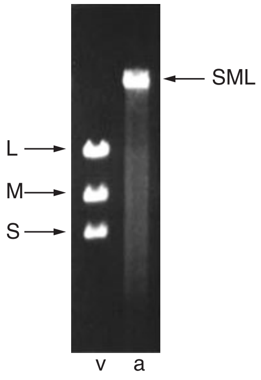


Figure 16-8 Agarose gel electrophoresis of double-stranded RNA isolated from virions. Lane v shows the distribution of normal segments L, M, and S. Lane a shows double-stranded RNA from ϕ2515, which contains the entire genome of ϕ6 in one segment. The migration of the RNA indicates a size of about 14,000 bp.

that were chimeras of segments S and M. In the case where S was at the 5′ end, the chimeric molecule replaced both the S and M segments; however, when M was at the 5′ end, the virus required an independent segment S for viability (figure 16-7). It was also possible to prepare virus with a single genomic segment carrying all the genetic information of the normal genome but with the *pac* sequence of S (figure 16-8) (36).

ϕ6 can establish a stable carrier state in infected cells. If virus is prepared with a reporter gene such as resistance to kanamycin (*kan*) in segment M, one can isolate kanamycin-resistant colonies after infection (34). The frequency of forming carrier state infections is modified by mutations in gene 5 of ϕ6, so that mutants can be isolated that form carrier state in about 10% of infections. Carrier state colonies can be passaged on plates containing antibiotic. It has been observed that in some cases, after many passages, deletions appear in segment S. In some cases, the segment is lost entirely. This is an apparent anomaly because the packaging model requires that segment S be packaged before the others. When the sequence of segment L in these carrier state colonies was determined, it was found that a mutation had occurred in gene 1, which codes for the major structural protein of the inner core. Procapsids produced with the mutated form of P1 were found to package segments M and L without the participation of S. Apparently, the mutation resulted in the loss of accessibility to the S binding site in the empty procapsid, so that the system behaved as though S were already packaged (37).

The finding that amino acid changes in protein P1 can change the packaging program suggests that this protein is the major determinant of packaging specificity. This is supported by the finding that bound RNA can be crosslinked to P1 (42). Directed changes in the *pac* sequences of ϕ6 segments S and M have been prepared and some of these reduce packaging by as much as 10,000-fold. Suppressor mutations that can accommodate the mutant RNA have been isolated and they map in gene 1 (42). A number of mutants in gene 1 have been isolated that result in the loss of packaging activity. Inner cores can be assembled lacking proteins P7 or P2 (4, 14, 19). In neither case is packaging specificity altered; however, the particles missing P7 show reduced packaging activity. Although protein P4 is the motor for packaging, there is no evidence that it is involved in the specificity of packaging. The location of the RNA binding sites on P1 and the mechanism by which the binding program changes have not been determined.

Reverse Genetics

The first system for reverse genetics in ϕ6 made use of the in vitro packaging and replication system (32). The Bamford laboratory had found that nucleocapsids of ϕ6 could infect spheroplasts of the host although they could not infect normal bacteria because they lacked the attachment proteins and the ability to penetrate the outer layers of the cell. They found that removing the shell of protein P8 led to the loss of infectivity, but that purified P8 could be reapplied to the particles in the presence of calcium ions. Core particles that had packaged RNA transcripts of cDNA plasmids could be coated with P8 and successfully infect spheroplasts (33). In this manner it was possible to prepare virus with genomic deletions or additions, particularly with reporter groups such as kanamycin resistance or *lac*α (35).

Once virus had been prepared with stable, nonreverting deletions, it was possible to test for the ability of live virus to acquire plasmid transcripts as replacements for defective genomic segments. It was found that transcripts that contained the full sequence content of a genomic segment, even if bounded by nonviral sequence at the 5′ and 3′ ends, could result in precise acquisition of the viral sequence (28). In cases where the transcript did not contain the viral 3′ sequences, they could still be acquired; however, in these cases the acquisition involved RNA recombination so that the transcript could acquire a copy of a 3′ end from one of the other genomic segments. This process is called heterologous or nonhomologous recombination (26). In the case of ϕ8, it was found that transcripts that did not contain the 5′ *pac* sequences could be acquired through a process of homologous recombination (38).

But the easiest and most efficient procedure for reverse genetics was found to be electroporation of cDNA plasmids of the three genomic segments into cells that contain either SP6 polymerase or T7 RNA polymerase (48). In these cases the transcripts, using SP6 or T7 promoters, start at the beginning of the viral sequences but the 3′ ends are bounded by nonviral sequence. The resulting phage has correct 5′ and 3′ terminal sequences. Using this technique it has been possible to prepare mutant forms of bacteriophages ϕ6, ϕ 8, and ϕ13. For ϕ6 and ϕ13, where the 5′ sequence begins with GG, T7 polymerase is preferred for the in vivo transcription, while for ϕ8, which begins with GA, SP6 polymerase is preferred although T7 polymerase works at a lower efficiency.

RNA-Dependent RNA Polymerase

The RNA polymerases of the *Cystoviridae* are part of the inner core particle. There seem to be 12 molecules per virion and they are probably located under the 5-fold vertices. The polymerases have some ill-defined sequence requirements for template usage, but they do not determine the specificity of packaging (24, 30, 53). In ϕ6 there is a difference in (+) strand synthesis between segment L and the other two segments. This is due to a sequence difference at the 5′ end of the (+) strand in that segment L has GU while S and M have GG. The S and M templates support (+) strand synthesis at about 10 times the rate of L and this is corrected by changing the 5′ terminal sequence of L to GG (8). This difference might be important for the temporal control of expression, in that the genes for segments S and M are expressed late in infection when their transcription is much greater than that of L. However, bacteriophage ϕ8 has no difference in the 5′ base sequence between the segments. In vitro, ϕ8 transcription is about equal for all three segments, while that for ϕ6 is heavily biased toward S and M.

The polymerase of ϕ6 has been purified and crystallized (3). The crystals were of a high quality and an atomic structure of the molecule was obtained through X-ray diffraction studies with 0.2 nm resolution (figure 16-9) (2). This was the first atomic structure for a polymerase of a double-stranded (ds) RNA virus. The structure is of special interest in several regards. It is remarkably similar to that of the hepatitis C virus, which is quite similar to that of poliovirus. Structures of complexes with nucleotides and oligonucleotides were also obtained. These structures have led to a model for the initiation of (+) and (−) strand synthesis (2). The polymerase has separate channels for substrate nucleotide triphosphates and for template. The polymerase can synthesize (+) strands from a dsRNA substrate with the concomitant unwinding and exclusion of the parental (+) strand, which does not enter the polymerase. It can also synthesize (−) strands on the template of single-stranded (+) strand RNA. In both cases the dsRNA leaves the polymerase through a channel that is normally blocked by the C-terminal domain of the polymerase. The initiation of polymerization involves the pairing of an NTP, usually GTP, with the second nucleotide from the 3′ end of the template. The template then backs up so that another nucleotide can pair with the first nucleotide from the 3′ end of the template. Polymerization then begins and continues with the extrusion of the dsRNA product from the polymerase.

The polymerase in ϕ6 procapsids does not begin RNA synthesis until the packaging of all three segments is complete. There seems to be control over the activity of the polymerase. However, purified polymerase is able to carry out (−) strand and (+) strand synthesis on various templates (53). The polymerase does not have great specificity in terms of the templates that it will use. The primary instrument for specificity in these viruses is the packaging mechanism. However, the polymerase does distinguish between the dsRNA substrates for transcription. In most members of the *Cystoviridae*, there is a difference in the second nucleotide of the 3′ end of the (−) strand. In ϕ6 the sequence of the (−) strand is 3′-CCUUU in segments S and M, while it is 3′-CAUUU in segment L. This difference results in transcription activity for M and S being 10–20 times that of L. This difference is probably important in the regulation of the temporal expression of the genes of the various segments. Those of segments S and M are expressed at late times in infection, while those of L are expressed early. The distinction between early and late gene expression seems to be a combination of the message stability of L and the overproduction of S and M. Apparently, at early times the S and M messages are effectively degraded so that they are not translated. Bacteriophage ϕ8 is the only member of the family that does not have a sequence difference at the 5′ end of the (+) strand between segment L and the others. It seems that ϕ8 uses a completely different method of temporal control.

Another problem for the polymerase is the choice between the 3′ of the (+) or the (−) strand in dsRNA for the transcription template. The polymerase normally chooses the 3′ end of the (−) strand because this sequence is very

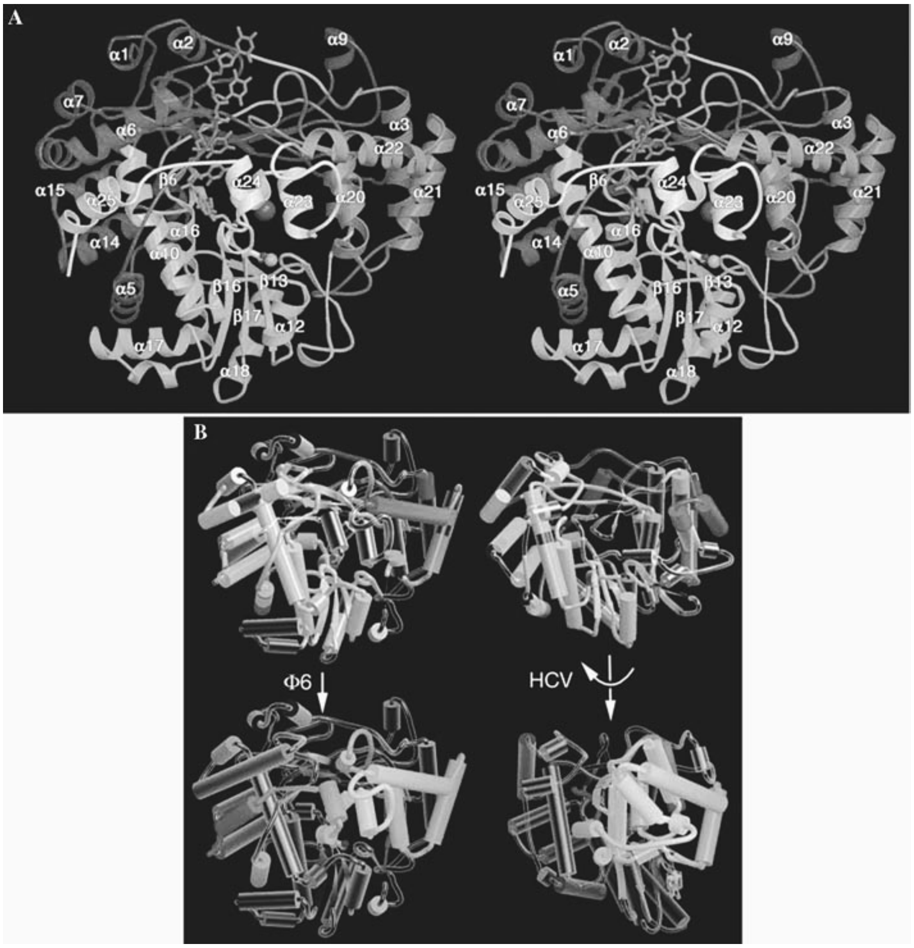


Figure 16-9 The structure of ϕ6 polymerase and comparison with hepatitis C virus (HCV) polymerase. A: Stereo image showing secondary structural elements. B: Comparison of ϕ6 and HCV polymerases (2). See thebacteriophages.org/frame_0160.htm for a color version of this figure.

high in A and U in all the members of the family. The ends should be able to melt out very easily. It has also been suggested that the polymerase prefers the sequence of (−) strand 3′ end for reasons apart from its low melting behavior (53).

Recombination

Recombination of RNA molecules has been demonstrated in almost all viral RNA systems. The mechanism of RNA recombination in viruses is template switching (23). The nascent chain detaches from its original template and then associates with a new template along with the original polymerase molecule or with a new polymerase. Recombination frequency is increased when one of the packaged (+) strands lacks a proper 3′ sequence and consequently cannot serve as a template for normal (−) strand synthesis (43). Most recombination seen in the *Cystoviridae* is heterologous in that the nascent chain seeks out a template with only about 3 nucleotides of identical sequence (26). However, ϕ8 is also capable of homologous recombination wherein the recombination occurs within a region of extensive sequence identity. Homologous recombination in that case requires about 600 nucleotides of identical sequence (38). It seems likely that heterologous recombination is more useful to these viruses than homologous recombination. The probability of packaging two molecules of the same segment class is very low, so homologous recombination would be a very inefficient means of correcting defective sequence. In these viruses, reassortment of segments occurs at high frequency and would be an effective means of sequence correction or change (5). The value of heterologous recombination would be in the acquisition of completely new sequence either from other viruses or from the host cell transcripts. In addition, it is a rather efficient means of correcting 3′ terminal deletions, which might be rather significant due to cellular nucleases. Packaging is solely dependent upon the integrity of the 5′ end, so it is possible to package molecules with defects at the 3′ end; these could be corrected by template switching from one of the other genomic segments. Although most of the evidence for recombination in the *Cystoviridae* is the result of experimental manipulation, the sequence of the 3′ end of segment M in ϕ13 is striking in that it is an exact copy of the 3′ end of segment M of ϕ6 although the rest of the segment has little sequence similarity (46). This sequence is also very different from that of the other two segments of ϕ13. It seems likely that the ϕ13 M terminal sequence is the result of a recombination event.

Sequence Differences Among the *Cystoviridae*

The members of the *Cystoviridae* are very similar in structure, but there are significant differences between many of them in terms of nucleotide and amino acid sequence as well as in some structural aspects (figure 16-10). The genomic segments are of approximately the same size in the members of the family. The sizes for ϕ6 are 2.9, 4.1, and 6.4 kbp respectively for S, M, and L. The arrangement of genes is rather similar as well. Segment L encodes the proteins of the inner core: P1, P2, P4, and P7. In ϕ13 gene *7* is the first gene in segment L; in ϕ6 there is a gene preceding gene *7* that is not essential for phage development. It is designated as gene *14*. Some of the very close relatives of ϕ6 such as ϕ10 have the same arrangement, while others such as ϕ7 and ϕ9 have an additional gene, designated as *orfE*, preceding gene *14*. These genes do not perform an essential function but may be involved in regulating the expression of gene *7* in a nonessential manner. ϕ12, which is rather distant from ϕ6, has two genes preceding gene *7*. ϕ8, which is the most distant member of the family, has a gene *14*. However, ϕ8 differs from all the members of the family in that the ortholog of gene *7* is found not near the 5′ end of L but rather at the 3′ end of segment L. The gene in the usual position of *7* has been renamed as gene *H* (49). The protein coded by gene *H* is not a component of the inner core, but does seem to play a significant role in the virus life cycle. Deletion of gene *H* results in a virus that forms very small, turbid plaques; however, mutants appear that are able to form plaques of almost normal size. It is not clear what role this protein plays.

ϕ7, ϕ9, and ϕ10 have very similar sequences to those of ϕ6. Base sequence identity is very high with most changes in the third base of codons. Phages ϕ8, ϕ12, and ϕ13 are very different from the others and from themselves in base sequence and even in amino acid sequences. ϕ13 is the closest of the three to ϕ6, yet the amount of amino acid sequence identity is 50% for the polymerase P2, while P1, P4, and P7 show about 30% identity. These figures can be contrasted with those for ϕ8 where there is almost no similarity in proteins P1, P4, and P7 and the polymerase shows only 20% identity with the polymerase of ϕ6.

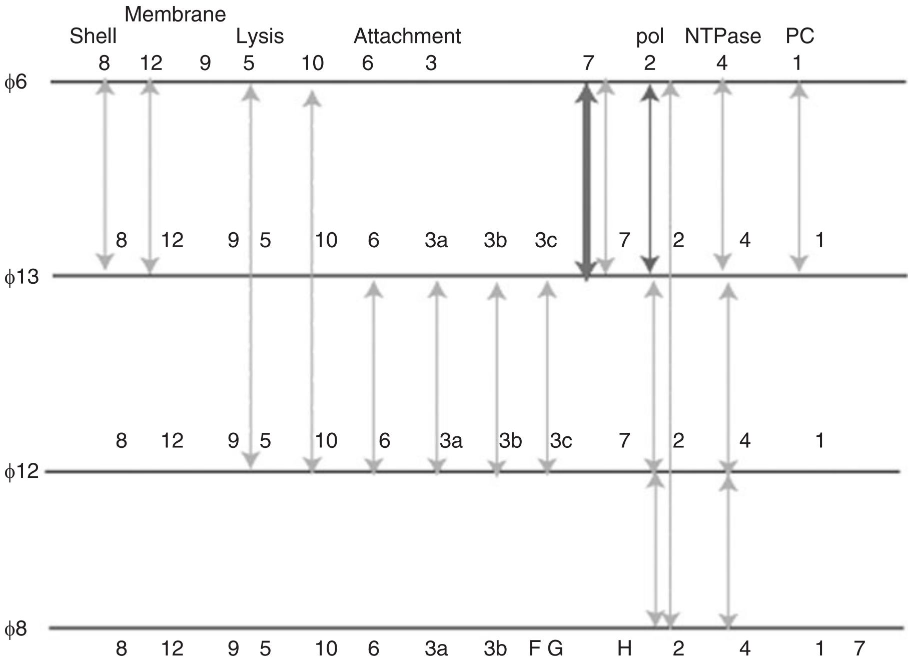

Membrane
Shell
Lysis
Attachment
pol
NTPase
PC
8 12 9 5 10 6 3 7 2 4 1
ϕ6
8 12 9 5 10 6 3a 3b 3c 7 2 4 1
ϕ13
8 12 9 5 10 6 3a 3b 3c 7 2 4 1
ϕ12
ϕ8
8 12 9 5 10 6 3a 3b F G H 2 4 1 7

Figure 16-10 Similarities in cystovirus genes. ϕ8 and ϕ12 share about 20% amino acid identity in P2, while ϕ13 has 50%. ϕ8 shares some sequence similarity in gene *4* with ϕ12, but nothing else. ϕ12 shows some similarity in the host attachment proteins to ϕ13 as well as similarity in gene *4*. ϕ12 also shows similarity in the lysis genes *5* and *10* with the ϕ6 family. ϕ13 has 30% amino acid identity with ϕ6 in the inner core protein genes and genes *8* and *12*, and a striking identity in base sequence for the 3′ end of segment M. See thebacteriophages.org/frames_0160.htm for a color version of this figure.

ϕ12 polymerase shows 21%, 24%, and 20% identity to the polymerases of ϕ6, ϕ8 and ϕ13. Protein P4 of ϕ12 shows 27% sequence identity to P4 of ϕ8, and ϕ13 (11).

There are two major structural distinctions within the family *Cystoviridae*. The first is that several of the viruses attach to the host LPS rather than to a type IV pilus. Whereas ϕ6 has an attachment apparatus consisting of the polytopic membrane protein P6 that holds the attachment specificity protein P3, ϕ8, ϕ12, and ϕ13 have a more complex attachment apparatus consisting of several P3 proteins such as P3a, P3b or P3c. The second distinction is that all the phages except ϕ8 have an intermediate shell of protein P8 that covers the inner core and is involved in membrane acquisition and host cytoplasmic membrane penetration. In ϕ8, the protein P8 is a component of the viral membrane and the inner core is able to penetrate the host cytoplasmic membrane on its own and it is also able to acquire the viral membrane directly.

Although the attachment proteins are completely different in ϕ6 and ϕ13, it is possible to substitute the M segment of ϕ6 for that of ϕ13 (46). Even though the *pac* sequences are completely different, the *pac* sequence of ϕ6 is able to be recognized by ϕ13 and to be maintained in a stable fashion. ϕ6 will not accept the *pac* sequence of ϕ13; however, the genes of ϕ13 are acceptable if the genomic segment is supplied with the *pac* sequence of ϕ6. The resulting membranes are an interesting mélange of the two systems. Protein P9 is the major membrane protein and is coded by segment S, protein P10 is an important component of the membrane and it would be coded by segment M along with P6 and the P3 assortment. Membrane assembly is determined primarily by protein P9 and the morphogenetic protein P12 (18). This membrane would have to interact with the shell of P8. These three genes are on segment S. The question is how the attachment apparatus of P6 and P3 recognizes the membrane composed of a heterologous P9. In any case, when ϕ6 uses the attachment proteins of ϕ13, it does not attach any more to pilus, but does attach to rough LPS.

ϕ13 shows 28%, 22%, and 22% identity to the ϕ6 proteins P6, P8, and P12 respectively. ϕ12 shows significant identity to the proteins P6 and P3a, b, and c of ϕ13. These are the host attachment proteins. The lysis proteins, P5 and P10, of ϕ12 show 30% and 62% identity to proteins of ϕ6.

It is striking that ϕ12 has acquired or maintained the lysis cassette that is similar to that of ϕ6 even though the two genes are on separate chromosomes. It is also striking that ϕ12 has maintained the attachment gene set that is typical of the LPS binding group (15). We have found that ϕ12 can acquire the attachment genes of ϕ6. It might be that ϕ12 originally had both the attachment and lysis genes of ϕ6. It seems apparent that there has been, and probably still is, considerable genetic interaction between the members of the *Cystoviridae*. The most dramatic example is the complete identity of the 3′ end of ϕ13M to that of ϕ6.

The stringency of the genomic packaging specificity serves to maintain the high efficiency of plating of the *Cystoviridae* but, at very low frequencies, the viruses are able to package RNA with imperfect *pac* sequences or no *pac* sequences at all (38). These RNA molecules cannot usually be maintained, but their sequences can be copied into the normal genomic segments by recombination. In this manner, these phages can exchange materials and acquire genetic information from unrelated sources.

The Relationship of the *Cystoviridae* to the *Reoviridae*

The members of the *Cystoviridae* have three double-stranded RNA genomic segments that are contained within a core structure composed of 120 copies of its major protein species. The reoviridae have 10, 11 or 12 double-stranded RNA genomic segments that are also contained within a core structure composed of 120 copies of their major protein species (16). The similarity in structure is striking, although the *Reoviridae* do not have a component like the P4 NTPase on the inner core particle. *Rotavirus* does have a nonstructural protein, NSP2, that may play a similar role; however, genomic packaging in the *Reoviridae* is not yet elucidated (50). As expected, the *Cystoviridae* transcripts are polycistronic whereas the transcripts of the *Reoviridae* are primarily monocistronic as found for most viruses of eukaryotes.

References

1. Butcher, S. J., T. Dokland, P. M. Ojala, D. H. Bamford, and S. D. Fuller. 1997. Intermediates in the assembly pathway of the double-stranded RNA virus ϕ6. EMBO J. 16:4477–4487.
2. Butcher, S. J., J. M. Grimes, E. V. Makeyev, D. H. Bamford, and D. I. Stuart. 2001. A mechanism for initiating RNA-dependent RNA polymerization. Nature 410:235–240.
3. Butcher, S. J., E. V. Makeyev, J. M. Grimes, D. I. Stuart, and D. H. Bamford. 2000. Crystallization and preliminary X-ray crystallographic studies on the bacteriophage ϕ6 RNA-dependent RNA polymerase. Acta. Crystallogr. 56:1473–1475.
4. Casini, G., X. Qiao, and L. Mindich. 1994. Reconstitution of active replicase in procapsids of the segmented dsRNA bacteriophage ϕ6. Virology 204:251–253.
5. Chao, L., T. T. Tran, and T. T. Tran. 1997. The advantage of sex in the RNA virus ϕ6. Genetics 147:953–959.
6. Day, L. A., and L. Mindich. 1980. The molecular weight of bacteriophage ϕ6 and its nucleocapsid. Virology 103:376–385.
7. de Haas, F., A. O. Paatero, L. Mindich, D. H. Bamford, and S. D. Fuller. 1999. A symmetry mismatch at the site of RNA packaging in the polymerase complex of dsRNA bacteriophage ϕ6. J. Mol. Biol. 294:357–372.
8. Dijk, A. A. V., M. Frilander, and D. H. Bamford. 1995. Differentiation between minus- and plus-strand

synthesis: polymerase activity of dsRNA bacteriophage ϕ6 in an in vitro packaging and replication system. Virology 211:320–323.

9. Frilander, M., and D. H. Bamford. 1995. In vitro packaging of the single-stranded RNA genomic precursors of the segmented double-stranded RNA bacteriophage ϕ6: the three segments modulate each other's packaging efficiency. J. Mol. Biol. 246:418–428.
10. Frilander, M., P. Gottlieb, J. Strassman, D. H. Bamford, and L. Mindich. 1992. Dependence of minus strand synthesis upon complete genomic packaging in the dsRNA bacteriophage ϕ6. J. Virol. 66:5013-5017.
11. Gottlieb, P., C. Potgieter, H. Wei, and I. Toporovsky. 2002. Characterization of ϕ12, a bacteriophage related to ϕ6: nucleotide sequence of the large double-stranded RNA (dsRNA). Virology 295:266–271.
12. Gottlieb, P., X. Qiao, J. Strassman, M. Frilander, and L. Mindich. 1994. Identification of the packaging regions within the genomic RNA segments of bacteriophage ϕ6. Virology 200:42–47.
13. Gottlieb, P., J. Strassman, and L. Mindich. 1992. Protein P4 of the bacteriophage ϕ6 procapsid has a nucleoside triphosphate-binding site with associated nucleoside triphosphate phosphohydrolase activity. J. Virol. 66:6220–6222.
14. Gottlieb, P., J. Strassman, X. Qiao, A. Frucht, and L. Mindich. 1990. In vitro replication, packaging and transcription of the segmented dsRNA genome of bacteriophage ϕ6: studies with procapsids assembled from plasmid encoded proteins. J. Bacteriol. 172:5774–5782.
15. Gottlieb, P., H. Wei, C. Potgieter, and I. Toporovsky. 2002. Characterization of ϕ12, a bacteriophage related to ϕ6: nucleotide sequence of the small and middle double-stranded RNA. Virology 293:118–124.
16. Grimes, J. M., J. N. Burroughs, P. Gouet, J. M. Diprose, R. Malby, S. Zientara, P. P. C. Mertens, and D. I. Stuart. 1998. The atomic structure of the bluetongue virus core. Nature 395:470–478.
17. Hoogstraten, D., X. Qioa, Y. Sun, A. Hu, S. Onodera, and L. Mindich. 2000. Characterization of ϕ8, a bacteriophage containing three double-stranded RNA genomic segments and distantly related to ϕ6. Virology 272:218–224.
18. Johnson, M. D., III, and L. Mindich. 1994. Plasmid directed assembly of the lipid-containing membrane of bacteriophage ϕ6. J. Bacteriol. 176:4124–4132.
19. Juuti, J. T., and D. H. Bamford. 1997. Protein P7 of phage ϕ6 RNA polymerase complex, acquiring RNA packaging activity by in vitro assembly of the purified protein onto deficient particles. J. Mol. Biol. 266:891–900.
20. Juuti, J. T., D. H. Bamford, R. Tuma, and G. J. Thomas Jr. 1998. Structure and NTPase activity of the RNA-translocating protein (P4) of bacteriophage ϕ6. J. Mol. Biol. 279:347–359.
21. Kainov, D. E., S. J. Butcher, D. H. Bamford, and R. Tuma. 2003. Conserved intermediates on the assembly pathway of double-stranded RNA bacteriophages. J. Mol. Biol. 328:791–804.
22. Kainov, D. E., M. Pirttimaa, R. Tuma, S. J. Butcher, G. J. Thomas, D. H. Bamford, and E. V. Makeyev. 2003. RNA packaging device of dsRNA bacteriophages: possibly as simple as hexamer of P4 protein. J. Biol. Chem., 278:48084–48091.
23. Kirkegaard, K., and D. Baltimore. 1986. The mechanism of RNA recombination in poliovirus. Cell 47:433–443.
24. Makeyev, E. V., and D. H. Bamford. 2000. Replicase activity of purified recombinant protein P2 of double-stranded RNA bacteriophage phi6. EMBO J. 19:124–131.
25. Mindich, L. 1988. Bacteriophage ϕ6: a unique virus having a lipid-containing membrane and a genome composed of three dsRNA segments, pp. 137–176. *In* K. Maramorosch, F. A. Murphy, and A. J. Shatkin (eds.) Advances in Virus Research, vol. 35. Academic Press, New York.
26. Mindich, L. 1996. Heterologous recombination in the segmented dsRNA genome of bacteriophage F6. Semin. Virol. 7:389–397.
27. Mindich, L. 1999. Precise packaging of the three genomic segments of the double-stranded-RNA bacteriophage ϕ6. Microbiol. Mol. Biol. Rev. 63:149–160.
28. Mindich, L. 1999. Reverse genetics of the dsRNA bacteriophage ϕ6. Adv. in Virus Res. 53:341–353.
29. Mindich, L., and D. H. Bamford. 1988. Lipid-containing bacteriophages, pp. 475–520. *In* R. Calendar (ed.) The Bacteriophages, vol. 2. Plenum Publishing, New York.
30. Mindich, L., X. Qiao, S. Onodera, P. Gottlieb, and M. Frilander. 1994. RNA structural requirements for stability and minus strand synthesis in the dsRNA bacteriophage ϕ6. Virology 202:258–263.
31. Mindich, L., X. Qiao, J. Qiao, S. Onodera, M. Romantschuk, and D. Hoogstraten. 1999. Isolation of additional bacteriophages with genomes of segmented double-stranded RNA. J. Bacteriol. 181:4505–4508.
32. Olkkonen, V. M., P. Gottlieb, J. Strassman, X. Qiao, D. H. Bamford, and L. Mindich. 1990. In vitro assembly of infectious nucleocapsids of bacteriophage ϕ6: formation of a recombinant double-stranded RNA virus. Proc. Natl. Acad. Sci. USA 87:9173–9177.
33. Olkkonen, V. M., P. Ojala, and D. H. Bamford. 1991. Generation of infectious nucleocapsids by in vitro assembly of the shell protein onto the polymerase complex of the dsRNA bacteriophage ϕ6. J. Mol. Biol. 218:569–581.
34. Onodera, S., V. M. Olkkonen, P. Gottlieb, J. Strassman, X. Qiao, D. H. Bamford, and L. Mindich. 1992. Construction of a transducing virus from dsRNA bacteriophage ϕ6: establishment of carrier states in host cells. J. Virol. 66:190–196.
35. Onodera, S., X. Qiao, P. Gottlieb, J. Strassman, M. Frilander, and L. Mindich. 1993. RNA structure and heterologous recombination in the dsRNA bacteriophage ϕ6. J. Virol. 67:4914–4922.
36. Onodera, S., X. Qiao, J. Qiao, and L. Mindich. 1998. Directed changes in the number of dsRNA genomic segments in bacteriophage ϕ6. Proc. Natl. Acad. Sci. USA 95:3920–3924.
37. Onodera, S., X. Qiao, J. Qiao, and L. Mindich. 1998. Isolation of a mutant that changes genomic packaging specificity in ϕ6. Virology 252:438–442.
38. Onodera, S., Y. Sun, and L. Mindich. 2001. Reverse genetics and recombination in ϕ8, a dsRNA bacteriophage. Virology 286:113–118.

39. Pirttimaa, M. J., and D. H. Bamford. 2000. RNA secondary structures of the bacteriophage phi6 packaging regions. RNA 6:880–889.
40. Poranen, M. M., A. O. Paatero, R. Tuma, and D. H. Bamford. 2001. Self-assembly of a viral molecular machine from purified protein and RNA constituents. Mol. Cell 7:845–854.
41. Qiao, X., G. Casini, J. Qiao, and L. Mindich. 1995. In vitro packaging of individual genomic segments of bacteriophage F6 RNA: serial dependence relationships. J. Virol. 69:2926–2931.
42. Qiao, X., J. Qiao, and L. Mindich. 2003. Analysis of the specific binding involved in genomic packaging of the dsRNA bacteriophage ϕ6. J. Bacteriol. 185:6409–6414.
43. Qiao, X., J. Qiao, and L. Mindich. 1997. An in vitro system for the investigation of heterologous RNA recombination. Virology 227:103–110.
44. Qiao, X., J. Qiao, and L. Mindich. 1995. Interference of bacteriophage ϕ6 genomic RNA packaging by hairpin structures. J. Virol. 69:5502–5505.
45. Qiao, X., J. Qiao, and L. Mindich. 1997. Stoichiometric packaging of the three genomic segments of dsRNA bacteriophage ϕ6. Proc. Natl. Acad. Sci. USA 94:4074–4079.
46. Qiao, X., J. Qiao, S. Onodera, and L. Mindich. 2000. Characterization of ϕ13, a bacteriophage related to ϕ6 and containing three dsRNA genomic segments. Virology 275:218–224.
47. Semancik, J. S., A. K. Vidaver, and J. L. Van Etten. 1973. Characterization of a segmented double-helical RNA from bacteriophage ϕ6. J. Mol. Biol. 78:617–625.
48. Sun, Y., X. Qiao, and L. Mindich. 2004. Construction of carrier state viruses with partial genomes of the segmented dsRNA bacteriophages. Virology 319:274–279.
49. Sun, Y., X. Qiao, J. Qiao, S. Onodera, and L. Mindich. 2003. Unique properties of the inner core of bacteriophage phi8, a virus with a segmented dsRNA genome. Virology 308:354–361.
50. Taraporewala, Z., D. Chen, and J. T. Patton. 1999. Multimers formed by the rotavirus nonstructural protein NSP2 bind to RNA and have ucleoside triphosphatase activity. J. Virol. 73:9934–9943.
51. Tuma, R., J. K. Bamford, D. H. Bamford, and G. J. Thomas, Jr. 1999. Assembly dynamics of the nucleocapsid shell subunit (P8) of bacteriophage phi6. Biochemistry 38:15025–15033.
52. Vidaver, A. K., R. K. Koski, and J. L. Van Etten. 1973. Bacteriophage ϕ6: a lipid-containing virus of *Pseudomonas phaseolicola*. J. Virol. 11:799–805.
53. Yang, H., E. V. Makeyev, and D. H. Bamford. 2001. Comparison of polymerase subunits from double-stranded RNA bacteriophages. J. Virol. 75:11088–11095.

This page intentionally left blank

PART IV

INDIVIDUAL TAILED PHAGES

This page intentionally left blank

17

The T1-Like Bacteriophages

GREGORY J. GERMAN
RAJEEV MISRA
ANDREW M. KROPINSKI

The Dark Side of T1 Phage

Although T1 got its initial fame from its use, in 1943, by Salvador Luria and Max Delbrück in their landmark fluctuation test (64), an entirely different reason has resulted in its notoriety since then. This phage's resistance to desiccation results in its persistence for weeks in an aerosol which in turn created a real nightmare for laboratories engaged in research employing *E. coli* K-12 strains. It was not uncommon to hear cries from laboratory workers that their overnight bacterial cultures contained only floating debris. This was the result of an onslaught by T1. Laboratories were quarantined for days until the airborne count of T1 vanished. Bacterial geneticists have had T1-resistant *E. coli tonA* mutants since the 1940s, and as a result many laboratories protected against unwanted invasion of T1 by introducing null *tonA* alleles into their *E. coli* strain collection. Even today T1 maintains its bad reputation because it remains a problem, particularly in laboratories that only casually use *E. coli* strains for experiments involving cloning, protein overproduction, and analysis of genomic libraries. Because of this, almost all commercially available transformation-competent *E. coli* cells now contain a *tonA* mutation. Despite the knowledge of T1-mediated calamities, many laboratories continue to work with *E. coli* strains that are not *tonA*. Apparently, it often takes a large-scale disaster for laboratories to become believers in the incidental havoc T1 can wreak on innocent bacterial cultures. Despite its infamy, phage T1 is an interesting phage in its own right. In this chapter we consider the biology of the T1-like phages, concentrating on phage T1 and the related phage TLS.

What's in the Name?

Phage TLS recently went through a name change due to its perplexing history (37). Prior to 2001, it was known as U3 in several laboratories across North America and Australia. But there were clearly two distinct strains of U3, as is evident from analysis of phage-resistant mutants. The U3 strain used in the laboratories of Carl Schnaitman (5, 74) and Rajeev Misra (58, 104, 105) yielded phage-resistant mutants that concurrently displayed a classical "deep-rough" phenotype (hypersensitivity to antibiotics, dyes, and detergents) of lipopolysaccharide (LPS)–defective mutants.

A different U3 strain, which also requires LPS as its receptor, was used by Malcolm Casadaban's laboratory. In their analysis of LPS mutants, the U3^r phenotype did not always correlate with the hypersensitivity phenotype (94). The fact that the two strains of U3 had different receptor specificities was evident from biochemical analyses of mutant LPS (74, 94). A U3 strain used in Paigen and Beacham's laboratories appeared to be similar to that used by Casadaban's (78, 111). They, like Casadaban's laboratory, reported the requirement of a galactose residue in the LPS core for infection. Watson and Paigen also showed that their U3 strain is a small, tailless phage resembling ϕX174 in its physical properties, and, unlike our U3 strain, infected *E. coli* K-12 but not *E. coli* B and C strains (111). As a result, this U3 strain was classified as a member of the genus microvirus (single-stranded DNA genomes) in the viral family *Microviridae*.

Based on host range, receptor specificities, structural features, and genome composition, the U3 strain used in Carl Schnaitman and our (Misra) laboratories is indisputably a different phage (with a flexible tail (figure 17-1) and a double-stranded DNA genome) than the one used by the Casadaban, Paigen, and Picken laboratories. This obliged a name change for the U3 strain reported later in the literature. Consequently, the U3 strain used by us, and described here, was re-named TLS, for TolC- and LPS-specific phage (37).

What Is a T1-Like Phage?

Using a rigid phenetic approach, the International Council for the Taxonomy of Viruses (ICTV) classifies phage into

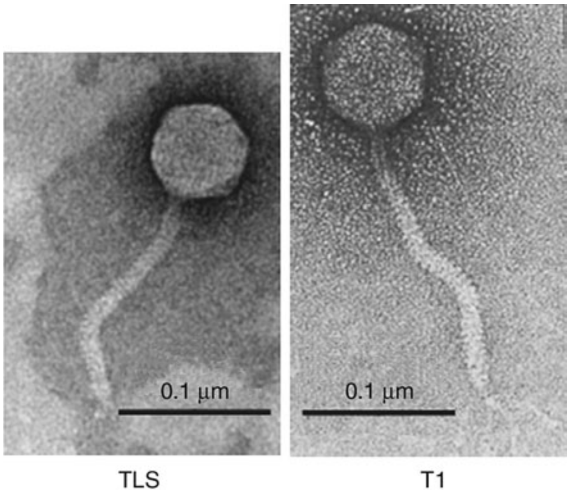


Figure 17-1 Electron micrographs of phage T1 (47; with permission) and phage TLS.

genera based upon analysis of ultrastructure, host range, lytic or temperate nature, genome size, and packaging peculiarities (2). On the basis of these criteria phage T1 has been classified as a lytic member of the *Siphoviridae* family (long noncontractile tails) specific for *Escherichia coli* and *Shigella* strains. Further distinguishing characteristics of T1 are an extremely flexible tail (151 nm in length by 8 nm in width) and an icosahedral head about 60 nm in diameter (figure 17-1). One complementary ICTV characteristic worth mentioning is the size difference in the two predominant head subunits between λ (38 and 53 kDa) and T1 (26 (suspected) and 33 kDa). The last characteristic of the T1-like phage genus is that the members have a genome less than 60 kb and have *pac* sites as opposed to *cos* sites. Headful packaging from a *pac* site creates genomes that are terminally redundant and partially circular permuted (81).

While there are over 50 phages that could be considered relatives of T1 based on morphological evidence, the ICTV has tentatively classified only the following enterobacterial phages into the T1-like group: 102, 103, 150, 168, 174, β4, D20, ϕγ, Hi, and UC-1. The TolC-specific phage TLS is under consideration by the ICTV for proper species status within this genus.

Another phenetic approach for classification as T1-similar ("similar" to delineate it from the T1-like genus description) phage could use a loose definition based on mutual recombination or packaging. This is in relation to the term "lambdoid," which is reserved for phages that can recombine with λ (e.g., HK97, P22, and HK620) regardless of their morphological characteristics or ICTV taxonomy. In a study by Hug and colleagues (47) a T1 type 3 variant lacking the ability to synthesize its own DNA was able to package other phages at different rates. The DNA from Hi, 150, 168, 172, and KD9 phages was packaged nearly as efficiently as T1 control DNA. Interestingly, D20 did not show efficient packaging even though it shares a highly similar pattern of virion proteins on denaturing polyacrylamide electrophoresis (SDS-PAGE) and, like phage 103, was the only other phage inactivated by anti-T1 serum (50).

The lack of sequence data has prevented taxonomic characterization using a phylogenetic approach like that seen for a highly conserved head subunit gene which created a T4 type group (109) or a proteomic comparison approach which depends on fully sequenced phage genomes (93).

T1 Adsorption and DNA Entry

The outer membrane components, including proteins and LPS, are often exploited by phages to gain entry into the bacterial cell. The early work by Theodore Puck and coworkers demonstrated that phage T1 requires 10^{-3} M divalent or 10^{-2} M monovalent ions for optimum attachment to its host, *E. coli* B. The adsorption rate constant ($\sim 3 \times 10^{-9}$ cm^3 min^{-1}) is extremely high, suggesting that most collisions lead to infection. While incubation at 37 °C leads to irreversible adsorption, incubation at 2 °C results in phages which are reversibly associated with the host cell. In order to investigate phage attachment further T1-resistant mutants of *E. coli* strain B were isolated and found to fall into two distinct classes. The first type referred to as *tonA* mutants (T one; lacking in strain B/1,5) do not allow any phage attachment, while the second class, *tonB*, permitted reversible but not irreversible adsorption (35, 79). The gene for the former phenotype, now referred to as *fhuA*, specifies a gated outer membrane (OM) protein involved in the transport of Fe^{3+}-ferrichrome. This protein also functions as the surface receptor for the phylogenetically unrelated phages (T1, T5, ϕ80, UC-1, and ES18) and the bacteriocins, colicin M and microcin J25 (12). The monomeric outer membrane protein FhuA has been crystallized (32) revealing a two-domain protein in which residues 160–714 form a β-barreled structure in the outer membrane with the N-terminal 159 residues forming a globular periplasmic plug or cork. The latter can be deleted with little change in the in vivo protein functions including ferrichrome uptake or phage attachment (11).

Random and site-directed mutagenesis of the surface-exposed FhuA peptide loops revealed that insertions in loop 4 (316–356) or loop 7 (502–515) influenced the phage sensitivity/resistance phenotype in an insertion-dependent manner (53). The insertion of four amino acid residues into FhuA loop 4 resulted in cells which were fully resistant to T1 and displayed reduced sensitivity to ϕ80. Insertions in loop 7 were fully resistant to colicin M. Subsequently, Killman and colleagues employed overlapping acylated hexapeptides in infection-inhibition assays demonstrating that the receptor specificities of the three phages overlapped to some extent, with T1 specifically recognizing residues ^{333}APADKGHY^{340} and ^{347}VDDEKLQ^{353} (52). Since ^{340}Y is involved in

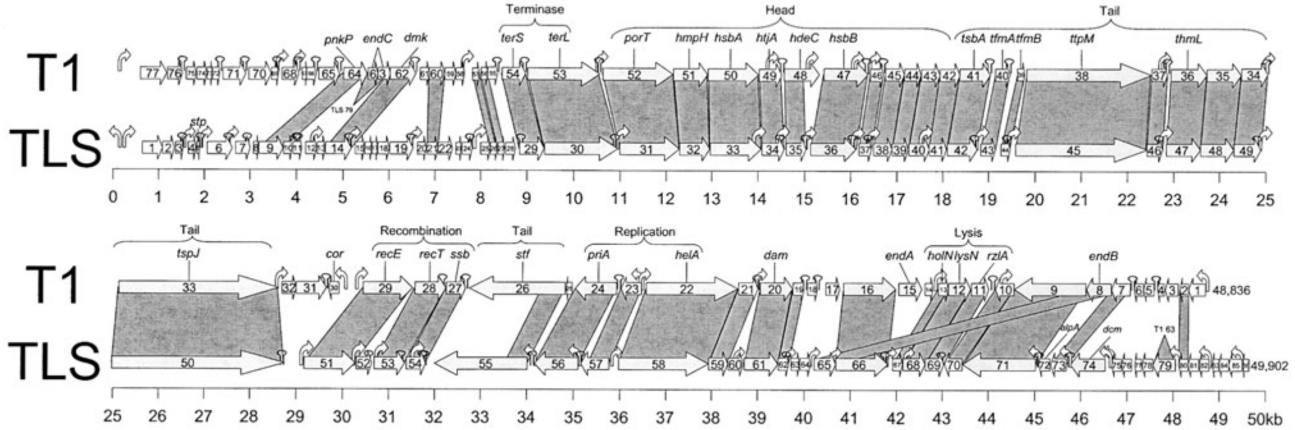


Figure 17-2 Genomic maps of T1 (above) and TLS (below).

ferrichrome binding these studies explain the old observation that ferrichrome can inhibit T1 adsorption (100).

Another T1-like phage, C1 (63), utilizes the OM vitamin B_{12} transport protein BtuB as its receptor, a feature it shares with phage BF23 (6) and the E group colicins and colicin R (61). Unlike T1, C1 phage infection is TonB-independent, requiring instead the inner membrane DcrA (SdaC) protein and the tethered-periplasmic protein DcrB. These are thought to initiate the creation of a DNA injection-competent channel (95).

UV-irradiated host cells allow reversible but not irreversible adsorption (35) indicating that infection is an energy-dependent process, a fact confirmed by the observation that TonB, which is involved in energy-translocating processes, interacts with FhuA in the periplasm (51). Phage T1 adsorption leads to the rapid loss of cellular ATP, K^+, and depolarization of the cell membrane ($\Delta\psi$). In the presence of high Mg^{2+} this phenomenon was transient, that is, the membrane pores became resealed (50). The DNA enters the cell in an oriented manner (39), which we interpret as being from the right-hand end of the genetic map presented in figure 17-2.

The TLS Receptor

Work carried out in Carl Schnaitman's laboratory reported that mutations in *tolC* and *rfa* (now called *waa*, 84) could confer resistance to phage TLS, indicating that both TolC and LPS are required for TLS infection (5). The *waaP* gene product, which phosphorylates the first heptose residue in the LPS core (98, 119), was found to be essential for TLS infection (74, 94, 104). The role of surface-exposed regions of TolC in TLS infection was recently confirmed through analysis of *tolC* missense mutants (37). Attempts to obtain mutations mapping in loci other than *tolC* and *waaP* have failed, thus reflecting the absence of any other nonessential cellular factors needed for TLS infection.

Because the number of LPS molecules on the bacterial cell surface is much greater than that of the TolC protein, it may be that LPS serves as an initial source of phage interaction. A reversible binding of TLS to LPS would then allow the phages to glide across the cell surface until they find TolC to initiate irreversible binding. If this were the case, overexpression of TolC in a LPS mutant (deleted for *waaP*) might circumvent the requirement for LPS. However, this was found not to be the case, showing that the requirement of LPS with its normal core is absolute. It also suggested that the role of LPS might not be simply to allow phages to find TolC but rather to serve as an essential component in the infection process.

Genomics

With the completion of the sequences of T1 (91) and TLS (38) many of the secrets of how these phages behave has been revealed. The unique sequence of the terminally redundant and circularly permuted T1 genome is 48,836 bp. Since the total mass of the virion DNA, based upon *Pst*I digestions, is 50.7 kb, the terminal repeats are 1.9 kb. This is less than the 2.8 kb previously suggested, and the genome is larger than the earlier estimates of 46.9–49.5 kb (27, 65, 89). In the case of TLS the unique sequence is 49,902 bp with 1 kb terminal repeats, for a total genome size of 50.9 kb. The overall base compositions of the two phages are 45.6 mol%G-C (T1 DNA) and 42.7 (TLS), which are both less than that of the host (51 mol%G-C). The A-T content profile exhibits two spikes: the first is located within the predicted left terminal repeat while the other is found downstream of the tail spike genes. The latter suggests possible lateral transfer of this specific region of the genomes. The difference in G-C content between the two phages is also reflected in the fact that there is remarkably little perfect sequence identity between the two genomes.

As with other phage genomes, the genes of these two phages are densely packed (figure 17-2). One unusual characteristic of the genes of these phages is the high percentage of small open reading frames (ORFs), which were particularly prevalent at the ends of the genome. The T1 genome contains 77 ORFs while that of TLS contains 86 (figure 17-2). Approximately one half of the phage proteins gave BLAST hits, but with few exceptions these showed a low degree of sequence similarity to their "homologs." The exceptions were the genes that we assume are involved in tail assembly, which showed >40% sequence identity, usually to corresponding proteins from coliphage N15 (83). The other gene showing a high degree of relatedness is that which encodes a lysozyme-like protein. The two phages shared 52 genes in common, with their proteins having 26–77% identical residues (average amino acid identity 53%).

Restriction and Modification

It has long been known that T1 DNA is insensitive to *Eco*BI [TGA(N8)TGCT] and *Eco*KI [AAC(N6)GTGC] type I restriction endonucleases, but is sensitive to *Eco*PI [AGACC] type III restriction endonuclease. In silico restriction analysis revealed that both T1 and TLS DNA lack *Bam*HI, *Kpn*I, *Sac*I, *Sac*II, or *Sph*I sites and have a statistical underrepresentation of other common restriction sites. In addition, T1 DNA does not contain *Eco*BI or *Eco*KI sites, answering the old question about how this phage escapes the common type I restriction endonucleases present in its hosts. On the other hand, phage TLS DNA has 13 *Eco*BI sites and a single *Eco*KI site.

The TLS gene 5 is highly similar to the *stp* gene found exclusively in T4 and T-even phages. T4's Stp is a 29 amino acid peptide which provides protection from host type I restriction enzymes including *Eco*PrrI. Stp is considered a "double-edge" sword during phage infection because it prevents DNA restriction at the expense of activating an endonuclease specific for the host tRNA$^{\mathrm{Leu}}$ anticodon (an ACNase) (77). Restriction of the anticodon furnishes an atypical cleavage that is difficult for the host to repair and leads to a nonproductive infection as well as sacrificial death of the cell. T4 infection circumvents the ACNase activity by producing Pnk (polynucleotide kinase) that reverses the unusual cleavage. *pnk* homologs are found in TLS (gene *9*) and T1 (gene *64*) genomes, albeit only the DNA coding for the equivalent C-terminal half of T4 Pnk that is responsible for phosphatase activity (33).

The only two genes that had been cloned from T1 were *dam* (*N*-6-adenine methyltransferase) and a small putative downstream gene called HP83 (99). The complete genomic sequences of T1 and TLS reveal that they both encode Dam methylases, whose significance is unknown since *E. coli* also expresses DNA adenine methylase activity. Other phages specific for Dam-positive bacteria such as coliphages T2, T4, and RB49 (26), *Haemophilus* phage HP2 (115), and *Shigella* phage SfV (4) also encode a copy of this apparently redundant protein. Lastly, TLS has a gene specifying a Dcm methylase, a feature which it shares with enterobacterial phages ε^{15} (NC_004775) and N15 (AAC19095). *E. coli* also possesses a methylase which adds a 5′-methyl group to the internal cytosine residues in CCAGG and CCTGG.

Transcription and Translation

Using pulse-labeling with ^{14}C-labeled amino acids Wagner et al. (110) demonstrated three classes of protein expression: early proteins (e.g., helicase, primase, and recombinase; 88), whose synthesis was shut off at the onset of DNA replication (108), an early-late class produced throughout the infective cycle, and a late class (e.g., structural proteins and lysozyme; 108) whose onset of synthesis was delayed. This suggests temporal regulation of late transcriptional or translational events, which was confirmed by an inability to detect late proteins in an in vitro coupled transcription/translation system (108).

Early transcription in coliphages belonging to the *Caudovirales* (1) normally involves host RNA polymerase recognition of promoter sites which are defined by the presence of the canonical hexamers (−35 TTGACA; −10 TATAAT) optimally separated by 15–19 bp (69). Transcript elongation terminates by rho-dependent or rho-independent mechanisms. Relatively little is known about transcription or temporal gene expression in T1. Unlike phages, such as those of the T7 group, which specify single-subunit RNA polymerases, transcription in T1-infected cells is fully dependent on the host RNA polymerase (109). Furthermore, neither T1 nor TLS contains genes homologous to RNA polymerase subunits. In the only study of T1 transcription, Gawron and colleagues (36) hybridized ^{32}P-labeled RNA, isolated early and late in phage infective cycle, to the alkali-denatured, electrophoretically separated strands of T1 DNA. They showed that while most of the hybridization occurred to the slower-migrating strand, at early times post-infection transcripts were observed to hybridize to the faster-migrating stand. Examination of the arrangement of genes on T1/TLS reveals that the bulk of the transcripts will be transcribed from the bottom strand, except early in infection when some transcript will come from the complementary strand. Early transcription probably involves divergent transcription of the helicase-*dam* and primase-tail fiber gene clusters resulting in DNA synthesis.

In both phages, the majority of putative late transcription would, upon first glance, appear to occur in a single block from left to right resulting in expression of the genes involved in packaging and morphogenesis. This modality is found, with minor variations, in all other phages. Closer examination reveals, particularly in the "late operon," that T1 and TLS both contain an unusually high content

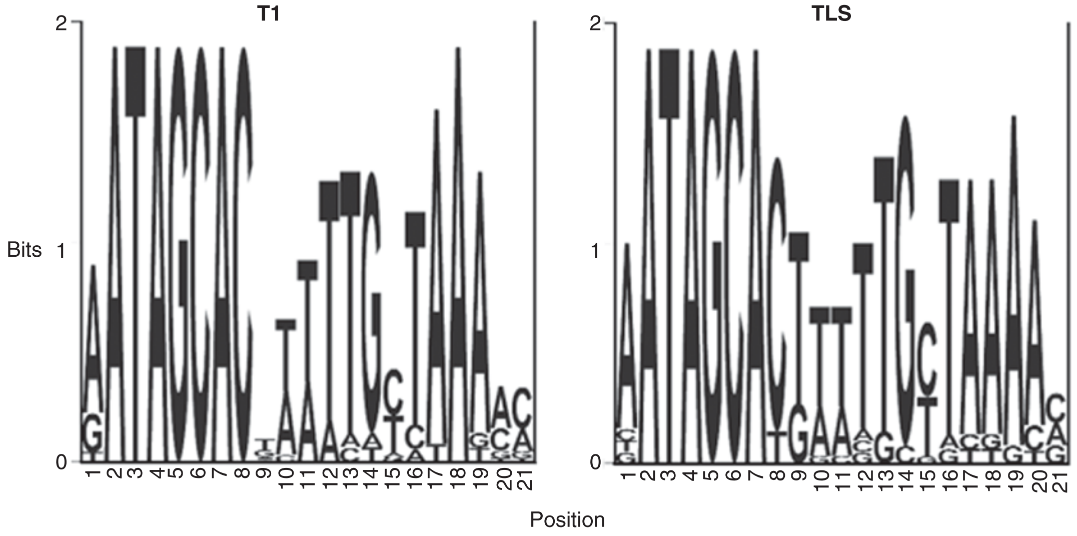


Figure 17-3 Consensus sequence logos of the T1 and TLS 21 nucleotide direct repeats. Produced using WebLogo (http://www.bio.cam.ac.uk/cgi-bin/seqlogo/logo.cgi).

of rho-independent terminators. In coliphage T4 these sites frequently contained a UUCG or GNRA loop sequence (70) that none of the putative T1/TLS terminators possess. The presence of transcriptional terminators led us to speculate that either a transcriptional read-through system exists such as that which occurs in lambdoid phage HK022 (48) or that downstream promoters must exist to direct transcription of subsequent genes. In both phages we have identified the latter (figure 17-2).

In addition both phages possess numerous 21 nucleotide direct repeats located in the intergenic regions or overlapping the translational terminators of the preceding gene. In the case of T1 we have identified 20, while TLS possesses 19. While their sequences (figure 17-3) are reminiscent of UP-elements in *E. coli* (54), their positions suggest that they may function in a manner equivalent to eukaryotic enhancers. This suggests a far different approach to transcriptional regulation than has been seen with other phages such as the two common modalities exemplified by the lambdoid phages and T7. In the former case, a single promoter regulates transcription of the morphogenesis genes, while with the latter phages multiple transcriptional start sites are located in the late region. TLS and T1 appear to divide the late region up into a series of transcriptional modules (transcriptons) flanked by rho-independent terminators and containing RpoD-dependent promoters and perhaps enhancers. This molecular approach may account for the short latent period of 13 minutes observed with coliphage T1 (9, 24, 90).

T1 infection causes a rapid cessation of host protein synthesis initially, presumably as a function of the effect of the phage of membrane function. But infection also has a lasting inhibitory affect on the translation of existing host mRNA (109). The reason for this is unknown.

Frameshifting

Programmed frameshifting is a phenomenon associated with translation, in which a stalled ribosome usually slips +1 or −1 and continues translating the message resulting in a protein with an altered C-terminus (3, 41). The slippery site AAA AAA GAG (LysLysGlu) occurs in *orf11*. This gene has a downstream stem-loop structure which would form a pseudoknot. Translational slippage would result in a slightly larger protein terminating in kheKKGGVSLLSYLYLLT rather than kheKKEA. In phage TLS another example of frameshifting occurs in *tfmA* specifying a minor tail protein, which results in a fusion with the downstream gene *tfmB*. A −1 frameshift at the slippery site GTA AAA AAC (ValLysAsn) results in a fusion protein VK(TfmA)-KKLKRAVYLYYQKPPTDAELQAVGLTRADYEGE DPPEVIFDES-(TfmB).

Host Genome Degradation and Phage DNA Replication/Recombination

In the following sections we will discuss the genetic organization of T1 from a modular perspective. T1 and TLS encode several proteins which play roles in DNA metabolism including its degradation, synthesis, and recombination. After phage infection the host DNA is degraded and approximately two thirds of its phosphorus content is recovered in progeny viruses (56, 57). In spite of this, host genome degradation is not essential for phage development (90). The mechanism of host DNA degradation is not understood. Interestingly, mutants in an early gene (previously designated 2.5) are able to block host DNA degradation without blocking phage DNA synthesis (89, 90).

How this functions to regulate phage or host degradative enzymes is not immediately apparent. Our analysis of the phage genomes reveals several genes potentially involved in nucleotide metabolism, including those having homology to two phage kinases (*pnk, dmk*). In the case of *pnk*, TLS and T1 homologs are labeled PnkP because the truncated proteins only match to the C-terminal domain of T4's Pnk that codes for the phosphatase domain.

Initiation of DNA replication involves the coordinated accumulation of replication-associated proteins at the origin of replication (Ori). In the case of *E. coli* this entails the initial binding of the replisome-organizer protein DnaA and the subsequent localized melting of the DNA duplex (16, 17). DnaA permits recruitment of DnaB (replicative DNA helicase) (101) or a complex of DnaB with the helicase-loading protein DnaC. Other proteins added to the replication complex include clamp-loading protein (DnaX; PolC proteins τ and γ), clamp-binding protein (DnaN), DnaG (DNA primase), single-strand binding protein (SSB), DNA gyrase, HU, integration host factor (IHF), and DNA polymerase III (PolC). The replication complex coordinates simultaneous synthesis on the leading and lagging strands (13). In the case of temperate phages P22 and λ, gpO functions as the replisome-organizer, recruiting to the origin of replication a loading factor (gpP) in the case of λ or gp*12* helicase with phage P22. GpP binds to both gpO and the host replication helicase (DnaB) resulting in enlargement of the replication bubble (117).

The work of Bourque and Christensen (10), employing host temperature-sensitive DNA replication mutants, showed that DNA polymerase III (PolC), primase (DnaG), and DNA clamp loader (DnaX) were required for T1 replication, while DnaA, DnaB, DnaC, and DnaT (primasomal protein i) were not. Other phage whose replication is independent of host DnaA protein include P22 (96), P4 (103), and SPP1 (75). There are conflicting reports on whether DnaG, the primase for the lagging strand, is required. DNA replication likely requires transcription for either torsional stress or priming of DNA, because rifampicin, a transcriptional inhibitor, also inhibits DNA replication.

Early T1 phage genetics and protein analysis showed that two phage genes, specifying proteins of 38 and 65 kDa, respectively, were required for DNA replication (88). Amber mutations in either of these genes gave a D0 phenotype, that is, they were defective in DNA synthesis. The 38 and 65 kDa proteins are probably equivalent to the products of the T1/TLS *priA* and *helA* genes, respectively. The presence of an ATP-dependent helicase explains the growth of T1 on *dnaB*^{ts} hosts (10). Helicases function to unwind double-stranded nucleotides in a 5′→3′ or 3′→5′ direction and are classified into five superfamilies of which T1/TLS HelA proteins are members of superfamily II (COG1061). Each of the proteins in these superfamilies contains up to seven identified motifs. In both phages, HelA protein Motif I (G^{62}KT^{64}) most probably corresponds to the Walker A box which is, along with the Walker B box (D^{148}ECH^{151}, Motif II), involved in MgATP interaction. Motif VI could be Q^{413}LLGRGMR^{420} (T1) or Q^{404}LLGRGMR^{411} (TLS), which bear more than passing resemblance to QTIGRAAR from UvrB (18). If this is correct then the terminal arginyl residue of this motif may interact with the gamma-phosphate of the bound ATP. The only problem faced with correlating Hel to the previously noted 65 kDa protein is that mutants deficient in the gene specifying the latter protein have been isolated. It is not apparent why it was possible to isolate T1 mutants deficient in what appears to be a "redundant" gene. An alternative hypothesis on the role of the putative helicase gene advanced by German et al. (38) is based on the sequence similarity between this protein and RecQ. The latter protein is the helicase involved with RecF-mediated Holliday junction formation (8).

Both phages also specify proteins (T1, 306 amino acids; TLS, 308 residues) with homology to prophage (CP-933I, ϕR73, Fels-1) and bacteriophage (P4) primases and, small single-stranded DNA binding proteins (SSB analogs). Our analysis and the results of Bourque and Christensen (10) suggest that leading strand synthesis involves PolC, DnaX, and DnaN proteins. The synthesis of the lagging DNA would, like that of T4, require SSB and a primosome containing presumably the phage helicase and primase proteins. The apparent requirement for both the host DNA primase (DnaG) and the phage homolog cannot be explained at this time. Therefore, it would appear that T1 replication is an example of a relatively simple DnaA-independent replication strategy which involves primase and helicase analogs but has additionally done away with the need for DnaC. Again, the presence of phage-encoded SSB homologs is quite common since they have been observed with phages such as T3, T4, T7, 933W, P1, and prophage CP-933V.

Using Grigoriev DNA skew analysis, the inflection, indicative of the replication origin, occurs within *helA* as it is in *Salmonella* phage P22. Phage replication origins are frequently characterized by iterons such as are found with r1t (123), A2 (71), BK5-T (67), ϕ31 (66), and TP901-1 (73). In neither phage T1 nor TLS is there evidence for iterons within *helA*.

Previous studies on phage T1 demonstrated that it encodes a general recombination system ("*grn*") composed of two genes (27, 85). While host RecABC will not substitute for T1 *grn*, the *E. coli* RecE (exonuclease VIII) recombinational pathway will (80). Both phages T1 and TLS specify homologs of the host RecE protein and for an Erf homolog. The latter is a member, along with host RecT, of a single-strand DNA annealing protein family which also functions in recombination. As with T7 replication there is no evidence for circular intermediates in T1 replication (81, 86, 87) and it is assumed that concatenated molecules, which are the substrate for headful packaging, are generated by end-to-end recombination. These are lacking in *grn* mutants leading to failure to produce packagable DNA.

Packaging

Conserved amongst all tailed bacteriophages is the two-subunit terminase complex which cuts DNA and packages it into proheads through the portal protein orifice. The smaller of the two terminase proteins is thought to recognize a packaging site, while the larger interacts with the portal protein, and catalyzes an ATP-dependent cleavage of the template (118). Depending on the terminase structure and phage DNA sequence, there are three common mechanisms used to package DNA into a phage head from the poly-genomic template called concatemers. These are unit-full (typified by λ or T7), random headful (characterized by T4), and *pac* site-initiated headful as observed with P22 (118).

T1-like phage terminases recognize a *pac* site on the concatemers to catalyze only the first round of packaging. Subsequent rounds of packaging are initiated at the original cut site with the finishing of subsequent rounds occurring somewhat randomly as the head fills with over one genomic length of DNA. This creates virion DNA with terminally redundant ends of approximately 1500–2200 bp in P22 (30, 97), 1000 bp for TLS (German and Misra, unpublished observations), and 2800 for T1 (65). Another corollary to headful packaging is that the DNA packaged in one virion has a different gene order compared with another virion packaged in a different round: this phenomenon is called circular permutation. With phage P22, T1, and TLS packaging does not result in an infinite number of permuted molecules because only three or four rounds of packaging occur on a given concatemer. The telltale symptom of *pac*-mediated packaging is that genomic DNA cut with restriction enzymes reveals submolar fragments resulting from only a portion of the mixed population (62). For instance, the TLS genome contains a 6 kb distinct submolar *Hin*dIII generated fragment which is created by *pac* cut near base 1 of the TLS genome and a *Hin*dIII site 6 kb downstream. There are also two submolar highly diffused bands at approximately 5 and 4 kb which presumably derive from a second and third packaging events and the common downstream *Hin*dIII site.

The P22 *pac* consensus site was initially proposed to be AAGATTA (20) and then modified through mutagenesis studies to GAAgATTTatCTGaaGT (118). The *pac* site for TLS is located within an 81 base stretch of DNA that does not contain cytosines in the top strand, and near the end of this stretch is a 5′AGATTT3′ sequence, where the first "A" is the first base of the TLS genome. Likewise a stretch of 60 bases without a cytosine in the top strand and a 5′AGATAT3′ at its end is found and the "G" would be the 225th base of the T1 genome (German, unpublished observations). This location matches well with previous studies that indicated the T1 *pac* site was approximately 1 kb upstream from one *Eco*RI site at 1323 bp (82).

Besides *pac* the T1 phage requires another DNA locus called *pip*. Overall packaging efficiency is reduced in T1 *pip* mutants and they are only able to catalyze a single round of packaging from templates (28). The *pip* phenotype is somewhat reminiscent of that observed with λ *cosQ* mutants, which are also only able to package the first event (114). In comparing the genetic and genomic maps, *pip* is located near *dam* (27). We have been unable to define the nucleotide sequence of *pip*.

Phage Assembly

The genetic map for T1 consisted of a block of 10 head genes (27), two of which likely encoded the small and large terminase subunits. TLS and T1 genomes have at least 13 genes from the small terminase subunit to just before the predicted tail subunit. In the "head region" their gene maps are considerably different from the eight prototypical patterns found in diverse genomes that are postulated to be a λ-supergroup of *Siphoviridae* (15). For instance, T1 and TLS appear to have the genes for two proteolytically processed head subunit homolog: subunit A (T1 gene *50* and TLS gene *33*) and subunit B (T1 gene *47* and TLS gene *36*). SDS-PAGE analysis (figure 17-4) of T1 and TLS showed a common abundant band at 27.9 kDa (for T1) or 29.1 kDa (for TLS). Mass spectroscopic analysis of T1's band indicated that it is a 26,588 Da processed product and is coded by T1 gene *47* (91). The T1 gene *47* corresponds to TLS gene *36* and they code for protein products of 35,290

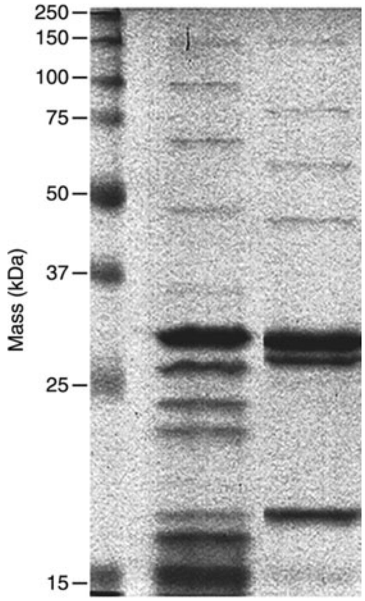


Figure 17-4 SDS-PAGE analysis of the structural proteins. Left: Phage T1. Right: Phage TLS. Approximately 10^{10} phage particles were loaded per lane. Gel was stained with Coomassie blue.

and 36,536 Da, respectively. The processing of the head subunit is reminiscent of HK97's where a 42 kDa precursor is processed to 31 kDa (29). Furthermore, the small size of the single gene between the portal gene and the first head subunit suggests that, like HK97, T1 gene *51* and TLS gene *32* code for a head protease. An extended 3′ portion found in T1 gene *48* versus TLS gene *35* reflects a difference between the genomes of two phages. This extended portion codes for an immunoglobulin-like domain, which is often found in phages that have head decoration proteins like T4's hoc (21). λ also encodes a head decoration protein (gpD) while P22, HK97, and L5 do not (19, 29).

The SDS-PAGE analysis of the two phages shows the presence of several additional bands for T1 (Figure 17-4). One band difference is possibly due to the T1 *48* gene product (calculated size 26,588 Da) while the corresponding gene *35* product for TLS (calculated size 16,355 Da) could comigrate with a putative FII-like head-tail joining protein found at a calculated 16,615 Da for TLS gp34. There is also a corresponding protein band at 17,007 Da for T1. The portal proteins of both phages are detected by SDS-PAGE analysis to be at their predicted locations of roughly 50 kDa (figure 17-4). Like the P22 portal protein (92), which is a 12mer, the T1 and TLS portal proteins appear to require cysteine cross-linkages because under nonreducing conditions the 50 kDa band is not detectable for either phage. T1 is also expected to carry an injectable phage protein in its head. This is based on observation that host protein synthesis was blocked at the level of messenger RNA translation before the commencement of any viral protein (109).

T1 was genetically shown to have at least nine genes responsible for producing phage tails (27). Eight of the nine were clustered after the head genes. The last gene was separated from the tail gene cassette by two recombination genes and thought to code for a tail assembly factor. The genomic map has shown that the tail assembly factor gene is actually the side tail fiber (*stf*) gene. Contrary to other known phage genomes, the transcription of the side tail fiber gene is in the direction opposite to the rest of the tail genes.

TLS's and T1's side tail fibers share strong homology for only the first 100 amino acids or roughly 20% of the protein. Thereafter TLS gp55 only shows homology with a *Salmonella typhi* prophage protein, while T1 gp26 shows exclusive homology with a hypothetical protein from *Photorhabdus luminescens*. Likewise, in the last published electron micrograph of T1, it was shown to have short and club-shaped tail fibers (47). Electron micrographs of TLS show long and thin tail fibers containing a kink (unpublished observations). T1-like phages contain an extremely flexible tail. The tail subunit, which is most likely the factor for the flexible tail, is not homologous to any other known phage but rather to prophage from *Yersinia*. The homology to known phages (HK97 and N15) starts with the tail tape measure gene and continues for the rest of the cassette, and therefore is likely to assemble in the prototypical pathway outlined for λ (19, 49). The last gene in the cassette, *tspJ*, codes for the predicted tail spike and host specificity factor. In λ, host range mutants were obtained by alterations localized at the C-terminus of the tail spike (112). Differences between T1 and TLS tail spike sequences are rather minor except that T1 contains a 30 amino acid repeat (starting at the 906th amino acid) preceding the region of λ host range mutants while TLS contains at least one insertion of 18 amino acids following this region (starting at the 1103rd amino acid).

Lysis

At the end of the latent period bacteriophage release is brought about through the concerted action of a holin which creates pores in the inner or cytoplasmic membrane, and an endolysin, which escapes through the pores to hydrolyze the peptidoglycan layer in the periplasm. In almost all cases these proteins are specified by a two-gene lysis cassette in which the holin gene precedes or overlaps the lysin gene (120, 121; chapter 10). Holins are characterized by their relatively small size (67–145 amino acid residues), usually contain two or three membrane-spanning helices, possess a charged C-terminus, and exhibit poor sequence identity to other members of this group of functionally similar proteins (40, 120, 121). The putative TLS and T1 holins contain predicted single-transmembrane domains and possess positively charged C-termini. Examples of other single-transmembrane domain holins include the *Borrelia burgdorferi* cp32 prophage holin protein BlyA (23), *Mycobacterium* phage Ms6 *hol* (34), and *Haemophilus influenzae* phage HP1 *hol* (31). Phage have evolved a variety of murein hydrolases and in the case of these two phages the endolysin genes specify a lysozyme which cleaves the β-1,4-linkages between adjacent *N*-acetylmuramic acid and *N*-acetylglucosamine residues in cell wall peptidoglycan.

In the case of both phages, downstream of the lysis cassette are located genes whose amino acid sequences reveal two transmembrane domains at both ends of the protein. No homologs have been discovered. Whether these putative gene products function as additional holins or are similar to the poorly defined lysis proteins Rz; 122 in phages such as λ and P22, is unknown.

Miscellaneous

T1 possesses three homing endonucleases of the HNH group, while TLS has two. These site-specific endonucleases are found in group I and group II introns (59) or as independent genes in bacteriophages (7, 19, 22). In all cases the proteins are homologous to *Xanthomonas oryzae* phage Xp10 (GenBank accession no. AY299121) and *Yersinia enterocolitica* phage ϕYeO3-12 (GenBank accession no. NC_001271) endonucleases. While phage Xp10 has 10 homing endonuclease genes, those from TLS and T1 are only homologous to four.

Morons are sequences, flanked by promoter and transcriptional terminators, often found inserted within a group of co-transcribed genes (48). Examination of the T1 genes (figure 17-2) reveals that *orf30* lies in the opposite orientation to its flanking genes and its orientation in coliphage N15. Furthermore, it is separated from *orf31* by a transcriptional terminator. This gene specified a protein called Cor involved in N15 lysogenic conversion gene, which is responsible for surface exclusion of T1, ϕ80, and N15 (83). Homologs are also synthesized by temperate phage HK022 (48) and ϕ80 (68). Why a virulent phage such as T1 should specify such a gene is unknown. TLS genes *52* and *60* are examples of genes inserted between mutually conserved regions of the two phages.

Evolution of the T1-Like Phages

Genomic analysis, predominantly on the temperate coliphages and viruses infecting bacteria involved in the dairy industry, has revealed the truth in the assertion that phage evolve through recombinational exchange with a large common genetic pool (14, 25, 43–45; chapter 4). The consequence of this is that many phage genomes are genetic mosaics with regions of homology interspersed with regions which have no homology. Access to this genetic pool will vary depending upon the ecological isolation of the host and the physiology of the phage. It has been argued that the virulent phages, particularly those that induce massive degradation of host DNA and have a well-organized replication or packaging strategies, will exhibit a lower incidence of horizontal gene transfer (55, 76). The high degree of homology between the essential genes of T1 and TLS and the strikingly similar overall genomic layouts in these two phages unequivocally confirm that they belong to the same lineage. Their genomes are also mosaics as revealed by the observation that the T1/TLS genomes contain regions with clear homology to other phage and prophage genomes. Homologs occur to phages infecting bacterial phyla *Actinobacteria* (mycobacteriophage Bxz1), Firmicutes (*Lactococcus* phage LL-H) and, as expected, members of the Proteobacteria. Two areas of particular interest are genes *38–30* (T1) and *45–50* (TLS) which are related to coliphage N15, and, with the exception of T1 gene *42*, genes *44–33* (TLS genes *39–50* except *41*), which are related to contiguous prophage sequences in *Yersinia pestis*.

One of the more interesting aspects of the genomics of T1/TLS is the presence of four linked genes which have been implicated in tail assembly in a number of members of the *Siphoviridae* infecting, or carried by, members of the class γ-Proteobacteria orders *Enterobacteriales* (*Salmonella* and *Escherichia*) or *Alteromonadales Shewanella*; (42). Phage ϕE125 resides in *Burkholderia thailandensis* (116), a member of the class β-Proteobacteria. A further unifying feature is that all the free-living phages (ϕE125, HK97, HK022, N15, and ϕ80) are classified as λ-like viruses at NCBI, suggesting that at a higher phylogenetic level T1/TLS might be considered to be part of a possible order "λ" within the *Siphoviridae*. Taking a "total evidence approach" to the origin of T1 a phylogenetic tree was constructed by alignment of "polyproteins" composed of the gp*37–34* and their homologs, with a *Homo sapiens* protein as the outlier. The results (figure 17-5) offer robust support for the existence of several lines of descent, which include λ and its prophage relatives (Gifsy-1, Gifsy-2, and Fels-1), the N15-HK97-HK022 cluster, and three deeply rooted clades involving T1/TLS, ϕE125, and λSo. The moron carrying Cor apparently was present in the phage genomes before the branching which gave rise to N15-HK022 and T1, and was lost in TLS and HK97.

It has been conservatively estimated that *E. coli* and *Salmonella enterica* diverged 120 million years ago (Myr) (60, 72). Employing the approach of Whittam et al. (113) we aligned common genes of *E. coli* K-12 with those of *S. enterica*

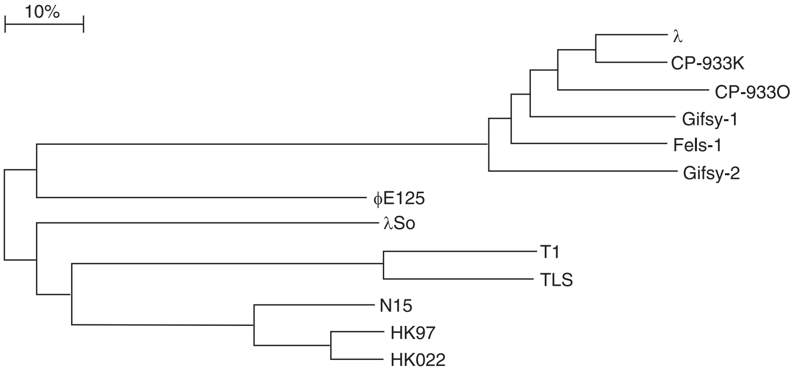


Figure 17-5 Phylogenetic analysis of tail cone assembly proteins using a "total evidence approach." The tree was constructed by alignment of "polyproteins" composed of the T1 gp*37–34*/TLS gp *46–49* and other homologs recovered from GenBank. The alignments were constructed using ClustalW (46) and analyzed further using TreeCon (106).

sv Typhimurium showing that these bacteria display an average of 18.3% divergence in the nucleotide sequence. From these values we calculated that a divergence of 1% is equivalent to 6.5 Myr. Since phages are obligate parasites it has been hypothesized that they evolve at a rate similar to their hosts (44). Using the phage *terS, terL, recT, hel*, portal, major head, and tail tape measure genes as chronometers of evolution, T1 and TLS exhibit 40.5% divergence. This is equivalent to 260 Myr. Using the same approach with the *terS*-major head gene cluster and the temperate lambdoid phages HK97 and HK022 (1.7% divergence), their last common ancestor was a mere 11 Myr ago. Two possibilities exist to explain these values: either the T1-like line of viruses is extremely old, or the rate of evolution of temperate phages is slower than that of their virulent cousins.

Unanswered Questions

While the completed genomic sequences of T1 and TLS have answered a number of important questions about these phages there are still areas in need of research.

1. Transcription: how transcription is temporally regulated; the number of promoters and the potential role of the 21 nucleotide direct repeats in transcription.
2. DNA replication: mechanism(s) of host DNA degradation; role of T1/TLS primase and helicase in phage DNA replication.
3. Modification: role of Dam and Dcm methylases.
4. Proteins: the cellular location and function of the large number of putative proteins showing transmembrane domains, and the correlation of T1 amber mutants with genes.
5. Tails: a hallmark of T1-like phages is that the tails are extremely flexible. Is the subunit geometry different from other phages? What is the nature of the receptor for the two tail spike proteins?
6. The mode of DNA entry.
7. A detailed study of morphogenesis of these phages and the location and role of the structural proteins.
8. Lysis: we have predicted that the holin has a single predicted transmembrane domain. The lysis cassette holin–endolysin–Rz homolog requires further analysis.
9. Stability: what physiochemical aspect of T1's structure contributes to the phage's extraordinary resilience to desiccation and persistence in aerosols?

Acknowledgments

R. M. would like to thank James DeGiulio for his assistance in sequencing and NIH grants GM48167 and GM066988. A. M. K. acknowledges the financial assistance of the Natural Sciences and Engineering Research Council of Canada and the contributions of Drs. Nancy Martin (proteomics), Pierre Lepage (sequencing), and Mary D. Roberts. We would also like to thank the *Journal of General Virology* for allowing us to reprint T1 in figure 17-1.

References

1. Ackermann, H.-W. 1999. Tailed bacteriophages: the order *Caudovirales*. Adv. Virus Res. 51:135–201.
2. Ackermann, H.-W., and M. S. DuBow. 2000. Description of virus taxa: Family *Siphoviridae*, pp. 85–89. *In* M. H. V. van Regenmortel, C. M. Fauquet, and D. H. L. Bishop (eds.) Virus Taxonomy: Classification and Nomenclature of Viruses. Academic Press, New York.
3. Alam, S. L., J. F. Atkins, and R. F. Gesteland. 1999. Programmed ribosomal frameshifting: much ado about knotting! Proc. Natl. Acad. Sci. USA 96:14177–14179.
4. Allison, G. E., D. Angeles, N. Tran-Dinh, and N. K. Verma. 2002. Complete genomic sequence of SfV, a serotype-converting temperate bacteriophage of *Shigella flexneri*. J. Bacteriol. 184:1974–1987.
5. Austin, E. A., J. F. Graves, L. A. Hite, C. T. Parker, and C. A. Schnaitman. 1990. Genetic analysis of lipopolysaccharide core biosynthesis by *Escherichia coli* K-12: insertion mutagenesis of the *rfa* locus. J. Bacteriol. 172:5312–5325.
6. Bassford, P. J., Jr., R. J. Kadner, and C. A. Schnaitman. 1977. Biosynthesis of the outer membrane receptor for vitamin B_{12}, E colicins, and bacteriophage BF23 by *Escherichia coli*: kinetics of phenotypic expression after the introduction of bfe+ and bfe alleles. J. Bacteriol. 129:265–275.
7. Belle, A., M. Landthaler, and D. A. Shub. 2002. Intronless homing: site-specific endonuclease *Seg*F of bacteriophage T4 mediates localized marker exclusion analogous to homing endonucleases of group I introns. Genes Dev. 16:351–362.
8. Bernstein, D. A., and J. L. Keck. 2003. Domain mapping of *Escherichia coli* RecQ defines the roles of conserved N- and C-terminal regions in the RecQ family. Nucleic Acids Res. 31:2778–2785.
9. Borchert, L. D., and H. Drexler. 1980. T1 genes which affect transduction. J. Virol. 33:1122–1128.
10. Bourque, L. W., and J. R. Christensen. 1980. The synthesis of coliphage T1 DNA: requirement for host DNA genes involved in elongation. Virology 102:310–316.
11. Braun, M., H. Killmann, and V. Braun. 1999. The beta-barrel domain of FhuADelta5-160 is sufficient for TonB-dependent FhuA activities of *Escherichia coli*. Mol. Microbiol. 33:1037–1049.
12. Braun, M., H. Killmann, E. Maier, R. Benz, and V. Braun. 2002. Diffusion through channel derivatives of the *Escherichia coli* FhuA transport protein. Eur. J. Biochem. 269:4948–4959.
13. Bruck, I., and M. O'Donnell. 2001. The ring-type polymerase sliding clamp family. Genome Biol. 2:reviews3001.1–reviews3001.3.
14. Brüssow, H. 2001. Phages of dairy bacteria. Annu. Rev. Microbiol. 55:283–303.

15. Canchaya, C., C. Proux, G. Fournous, A. Bruttin, and H. Brussow. 2003. Prophage genomics. Microbiol. Mol. Biol. Rev. 67:238–276.
16. Carr, K. M., and J. M. Kaguni. 2001. Stoichiometry of DnaA and DnaB protein in initiation at the *Escherichia coli* chromosomal origin. J. Biol. Chem. 276: 44919–44925.
17. Carr, K. M., and J. M. Kaguni. 2002. *Escherichia coli* DnaA protein loads a single DnaB helicase at a DnaA box hairpin. J. Biol. Chem. 277:39815–39822.
18. Caruthers, J. M., and D. B. McKay. 2002. Helicase structure and mechanism. Curr. Opin. Struct. Biol. 12:123–133.
19. Casjens, S., and R. Hendrix. 1988. Control mechanisms in dsDNA bacteriophage assembly, pp. 15–91. *In* R. Calendar (ed.) The Bacteriophages. Plenum Press, New York.
20. Casjens, S., W. M. Huang, M. Hayden, and R. Parr. 1987. Initiation of bacteriophage P22 DNA packaging series. Analysis of a mutant that alters the DNA target specificity of the packaging apparatus. J. Mol. Biol. 194: 411–422.
21. Childs, J. D., and R. Pilon. 1983. Evidence that bacteriophage T4 *eph1* is a missense *hoc* mutation. J. Virol. 46:629–631.
22. Dalgaard, J. Z., A. J. Klar, M. J. Moser, W. R. Holley, A. Chatterjee, and I. S. Mian. 1997. Statistical modeling and analysis of the LAGLIDADG family of site-specific endonucleases and identification of an intein that encodes a site-specific endonuclease of the HNH family. Nucleic Acids Res. 25:4626–4638.
23. Damman, C. J., C. H. Eggers, D. S. Samuels, and D. B. Oliver. 2000. Characterization of *Borrelia burgdorferi* BlyA and BlyB proteins: a prophage-encoded holin-like system. J. Bacteriol. 182:6791–6797.
24. Delbrück, M. 1945. The burst size distribution in the growth of bacterial viruses. J. Bacteriol. 50:131–135.
25. Desiere, F., W. M. McShan, D. van Sinderen, J. J. Ferretti, and H. Brussow. 2001. Comparative genomics reveals close genetic relationships between phages from dairy bacteria and pathogenic streptococci: evolutionary implications for prophage–host interactions. Virology 288:325–341.
26. Desplats, C., C. Dez, F. Tetart, H. Eleaume, and H. M. Krisch. 2002. Snapshot of the genome of the pseudo-T-even bacteriophage RB49. J. Bacteriol. 184:2789–2804.
27. Drexler, H. 1988. Bacteriophage T1, pp. 235–258. *In* R. Calendar (ed.) The Bacteriophages. Plenum Press, New York.
28. Drexler, H., and J. R. Christensen. 1986. T1 *pip*: a mutant which affects packaging initiation and processive packaging of T1 DNA. Virology 150:373–380.
29. Duda, R. L., K. Martincic, Z. Xie, and R. W. Hendrix. 1995. Bacteriophage HK97 head assembly. FEMS Microbiol. Rev. 17:41–46.
30. Duffy, C., and M. Feiss. 2002. The large subunit of bacteriophage lambda's terminase plays a role in DNA translocation and packaging termination. J. Mol. Biol. 316: 547–561.
31. Esposito, D., W. P. Fitzmaurice, R. C. Benjamin, S. D. Goodman, and J. J. Scocca. 1996. The complete nucleotide sequence of bacteriophage HP1 DNA. Nucleic Acids Res. 24:2360–2368.
32. Ferguson, A. D., J. Breed, K. Diederichs, W. Welte, and J. W. Coulton. 1998. An internal affinity-tag for purification and crystallization of the siderophore receptor FhuA, integral outer membrane protein from *Escherichia coli* K-12. Protein Sci. 7:1636–1638.
33. Galburt, E. A., J. Pelletier, G. Wilson, and B. L. Stoddard. 2002. Structure of a tRNA repair enzyme and molecular biology workhorse: T4 polynucleotide kinase. Structure 10:1249–1260.
34. Garcia, M., M. Pimentel, and J. Moniz-Pereira. 2002. Expression of mycobacteriophage Ms6 lysis genes is driven by two sigma(70)-like promoters and is dependent on a transcription termination signal present in the leader RNA. J. Bacteriol. 184:3034–3043.
35. Garen, A., and T. T. Puck. 1951. The first two steps of the invasion of host cells by bacterial viruses: II. J. Exp. Med. 94:177–189.
36. Gawron, M. C., J. R. Christensen, and T. M. Shoemaker. 1980. Exclusion of bacteriophage T1 by bacteriophage lambda. II. Synthesis of T1-specific macromolecules under N-mediated excluding conditions. J. Virol. 35:93–104.
37. German, G. J., and R. Misra. 2001. The TolC protein of *Escherichia coli* serves as a cell-surface receptor for the newly characterized TLS bacteriophage. J. Mol. Biol. 308:579–585.
38. German, G. J., J. DeGiulio, and R. Misra. 2005. The T1-like TolC- and lipopolysaccharide-specific (TLS) bacteriophage genome and the evolution of virulent phages. Submitted.
39. Gill, G. S., and L. A. MacHattie. 1975. Oriented extrusion of DNA from coliphage T1 particles. Virology 65:297–303.
40. Grundling, A., U. Blasi, and R. Young. 2000. Biochemical and genetic evidence for three transmembrane domains in the class I holin, lambda S. J. Biol. Chem. 275:769–776.
41. Harger, J. W., A. Meskauskas, and J. D. Dinman. 2002. An "integrated model" of programmed ribosomal frameshifting. Trends Biochem. Sci. 27:448–454.
42. Heidelberg, J. F., I. T. Paulsen, K. E. Nelson, E. J. Gaidos, W. C. Nelson, T. D. Read, J. A. Eisen, R. Seshadri, N. Ward, B. Methe, R. A. Clayton, T. Meyer, A. Tsapin, J. Scott, M. Beanan, L. Brinkac, S. Daugherty, R. T. DeBoy, R. J. Dodson, A. S. Durkin, D. H. Haft, J. F. Kolonay, R. Madupu, J. D. Peterson, L. A. Umayam, O. White, A. M. Wolf, J. Vamathevan, J. Weidman, M. Impraim, K. Lee, K. Berry, C. Lee, J. Mueller, H. Khouri, J. Gill, T. R. Utterback, L. A. McDonald, T. V. Feldblyum, H. O. Smith, J. C. Venter, K. H. Nealson, and C. M. Fraser. 2002. Genome sequence of the dissimilatory metal ion-reducing bacterium *Shewanella oneidensis*. Nature Biotechnol. 20:1118–1123.
43. Hendrix, R. W. 1999. Evolution: the long evolutionary reach of viruses. Curr. Biol. 9:R914–R917.
44. Hendrix, R. W. 2002. Bacteriophages: evolution of the majority. Theoret. Pop. Biol. 61:471–480.
45. Hendrix, R. W., M. C. Smith, R. N. Burns, M. E. Ford, and G. F. Hatfull. 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. Proc. Natl. Acad. Sci. USA 96:2192–2197.
46. Higgins, D. G., J. D. Thompson, and T. J. Gibson. 1996. Using CLUSTAL for multiple sequence alignments. Methods Enzymol. 266:383–402.
47. Hug, H., R. Hausmann, J. Liebeschuetz, and D. A. Ritchie. 1986. In vitro packaging of foreign DNA into heads of bacteriophage T1. J. Gen. Virol. 67:333–343.

48. Juhala, R. J., M. E. Ford, R. L. Duda, A. Youlton, G. F. Hatfull, and R. W. Hendrix. 2000. Genetic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. J. Mol. Biol. 299:27–51.
49. Katsura, I. 2003. Tail assembly and injection, pp. 331–346. *In* R. W. Hendrix, J. W. Roberts, F. W. Stahl, and R. A. Weisberg (eds.) Lambda II. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
50. Keweloh, H., and E. P. Bakker. 1984. Permeability changes in the cytoplasmic membrane of *Escherichia coli* K-12 early after infection with bacteriophage T1. J. Bacteriol. 160:347–353.
51. Killmann, H., C. Herrmann, A. Torun, G. Jung, and V. Braun. 2002. TonB of *Escherichia coli* activates FhuA through interaction with the beta-barrel. Microbiology 148:3497–3509.
52. Killmann, H., G. Videnov, G. Jung, H. Schwarz, and V. Braun. 1995. Identification of receptor binding sites by competitive peptide mapping: phages T1, T5, and phi 80 and colicin M bind to the gating loop of FhuA. J. Bacteriol. 177:694–698.
53. Koebnik, R., and V. Braun. 1993. Insertion derivatives containing segments of up to 16 amino acids identify surface- and periplasm-exposed regions of the FhuA outer membrane receptor of *Escherichia coli* K-12. J. Bacteriol. 175:826–839.
54. Kolasa, I. K., T. Lozinski, and K. L. Wierzchowski. 2002. Effect of A_n tracts within the UP element proximal subsite of a model promoter on kinetics of open complex formation by *Escherichia coli* RNA polymerase. Acta Biochim. Pol. 49:659–669.
55. Kovalyova, I. V., and A. M. Kropinski. 2003. The complete genomic sequence of lytic bacteriophage gh-1 infecting *Pseudomonas putida*: evidence for close relationship to the T7 group. Virology 311:305–315.
56. Labaw, L. W. 1951. The origin of phosphorus in *Escherichia coli* bacteriophages. J. Bacteriol. 62:169–173.
57. Labaw, L. W. 1953. The origin of phosphorus in the T1, T5, T6 and T7 bacteriophages of *Escherichia coli*. J. Bacteriol. 66:429–436.
58. Laird, M. W., A. W. Kloser, and R. Misra. 1994. Assembly of LamB and OmpF in deep rough lipopolysaccharide mutants of *Escherichia coli* K-12. J. Bacteriol. 176:2259–2264.
59. Landthaler, M., and D. A. Shub. 2003. The nicking homing endonuclease I-*Bas*I is encoded by a group I intron in the DNA polymerase gene of the *Bacillus thuringiensis* phage Bastille. Nucleic Acids Res. 31:3071–3077.
60. Lawrence, J. G., and H. Ochman. 1998. Molecular archaeology of the *Escherichia coli* genome. Proc. Natl. Acad. Sci. USA 95:9413–9417.
61. Lazdunski, C. J., E. Bouveret, A. Rigal, L. Journet, R. Lloubes, and H. Benedetti. 1998. Colicin import into *Escherichia coli* cells. J. Bacteriol. 180:4993–5002.
62. Le Marrec, C., D. van Sinderen, L. Walsh, E. Stanley, E. Vlegels, S. Moineau, P. Heinze, G. Fitzgerald, and B. Fayard. 1997. Two groups of bacteriophages infecting *Streptococcus thermophilus* can be distinguished on the basis of mode of packaging and genetic determinants for major structural proteins. Appl. Environ. Microbiol. 63:3246–3253.
63. Likhacheva, N. A., V. V. Samsonov, V. V. Samsonov, and S. P. Sineoky. 1996. Genetic control of the resistance to phage C1 of *Escherichia coli* K-12. J. Bacteriol. 178:5309–5315.
64. Luria, S., and M. Delbruck. 1943. Mutations of bacteria from virus sensitivity to virus resistance. Genetics 28:491.
65. MacHattie, L. A., M. Rhoades, and C. A. J. Thomas. 1972. Large repitition in the non-permuted nucleotide sequence of bacteriophage T1 DNA. J. Mol. Biol. 72:645–656.
66. Madsen, S. M., D. Mills, G. Djordjevic, H. Israelsen, and T. R. Klaenhammer. 2001. Analysis of the genetic switch and replication region of a P335-type bacteriophage with an obligate lytic lifestyle on *Lactococcus lactis*. Appl. Environ. Microbiol. 67:1128–1139.
67. Mahanivong, C., J. D. Boyce, B. E. Davidson, and A. J. Hillier. 2001. Sequence analysis and molecular characterization of the *Lactococcus lactis* temperate bacteriophage BK5-T. Appl. Environ. Microbiol. 67:3564–3576.
68. Matsumoto, M., N. Ichikawa, S. Tanaka, T. Morita, and A. Matsushiro. 1985. Molecular cloning of phi80 adsorption-inhibiting *cor* gene. Jpn. J. Genet. 60:483.
69. McLean, B. W., S. L. Wiseman, and A. M. Kropinski. 1997. Functional analysis of sigma-70 consensus promoters in *Pseudomonas aeruginosa* and *Escherichia coli*. Can. J. Microbiol. 43:981–985.
70. Miller, E. C., E. Kutter, G. Mosig, F. Arisaka, T. Kunisawa, and W. Rüger. 2003. Bacteriophage T4 genome. Microbiol. Mol. Biol. Rev. 67:86–156.
71. Moscoso, M., and J. E. Suarez. 2000. Characterization of the DNA replication module of bacteriophage A2 and use of its origin of replication as a defense against infection during milk fermentation by *Lactobacillus casei*. Virology 273:101–111.
72. Ochman, H., and A. C. Wilson. 1987. Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. J. Mol. Evol. 26:74–86. [Erratum appears in J. Mol. Evol. 1987; 26:377.]
73. Ostergaard, S., L. Brondsted, and F. K. Vogensen. 2001. Identification of a replication protein and repeats essential for DNA replication of the temperate lactococcal bacteriophage TP901-1. Appl. Environ. Microbiol. 67: 774–781.
74. Parker, C. T., A. W. Kloser, C. A. Schnaitman, M. A. Stein, S. Gottesman, and B. W. Gibson. 1992. Role of the *rfaG* and *rfaP* genes in determining the lipopolysaccharide core structure and cell surface properties of *Escherichia coli* K-12. J. Bacteriol. 174:2525–2538.
75. Pedre, X., F. Weise, S. Chai, G. Luder, and J. C. Alonso. 1994. Analysis of *cis* and *trans* acting elements required for the initiation of DNA replication in the *Bacillus subtilis* bacteriophage SPP1. J. Mol. Biol. 236:1324–1340.
76. Pedulla, M. L., M. E. Ford, J. M. Houtz, T. Karthikeyan, C. Wadsworth, J. A. Lewis, D. Jacobs-Sera, J. Falbo, J. Gross, N. R. Pannunzio, W. Brucker, V. Kumar, J. Kandasamy, L. Keenan, S. Bardarov, J. Kriakov, J. G. Lawrence, W. R. Jacobs, Jr., R. W. Hendrix, and G. F. Hatfull. 2003.

Origins of highly mosaic mycobacteriophage genomes. Cell 113: 171–182.

77. Penner, M., I. Morad, L. Snyder, and G. Kaufmann. 1995. Phage T4-coded Stp: double-edged effector of coupled DNA and tRNA-restriction systems. J. Mol. Biol. 249:857–868.

78. Picken, R. N., and I. R. Beacham. 1977. Bacteriophage-resistant mutants of *Escherichia coli* K12. Location of receptors within the lipopolysaccharide. J. Gen. Microbiol. 102:305–318.

79. Puck, T. T., A. Garen, and J. Cline. 1951. The first two steps of the invasion of host cells by bacterial viruses. I. The role of ions in the primary reaction. J. Exp. Med. 93:65–88.

80. Pugh, J. C., and D. A. Ritchie. 1984. Formation of phage T1 concatemers by the RecE recombination pathway of *Escherichia coli*. Virology 135:200–206.

81. Pugh, J. C., and D. A. Ritchie. 1984. The structure of replicating bacteriophage T1 DNA: comparison between wild type and DNA-arrest mutant infections. Virology 135:189–199.

82. Ramsay, N., and D. A. Ritchie. 1980. A physical map of the permuted genome of bacteriophage T1. Mol. Gen. Genet. 179:669–675.

83. Ravin, V., N. Ravin, S. Casjens, M. E. Ford, G. F. Hatfull, and R. W. Hendrix. 2000. Genomic sequence and analysis of the atypical temperate bacteriophage N15. J. Mol. Biol. 299:53–73.

84. Reeves, P. R., M. Hobbs, M. A. Valvano, M. Skurnik, C. Whitfield, D. Coplin, N. Kido, J. Klena, D. Maskell, C. R. Raetz, and P. D. Rick. 1996. Bacterial polysaccharide synthesis and gene nomenclature. Trends Microbiol. 4:495–503.

85. Ritchie, D. A., J. R. Christensen, J. C. Pugh, and L. W. Bourque. 1980. Genes of coliphage T1 whose products promote general recombination. Virology 105:371–378.

86. Ritchie, D. A., and D. H. Joicey. 1978. Formation of concatemeric DNA as an intermediate in the replication of bacteriophage T1 DNA molecules. J. Gen. Virol. 41:609–622.

87. Ritchie, D. A., and D. H. Joicey. 1980. Identification of some steps in the replication of bacteriophage T1 DNA. Virology 103:191–198.

88. Ritchie, D. A., D. H. Joicey, and D. T. Martin. 1983. Correlation of genetic loci and polypeptides specified by bacteriophage T1. J. Gen. Virol. 64:1355–1363.

89. Roberts, M. D. 2001. T1-like viruses, pp. 1–10. *In* C. A. Tidona and G. Darai (eds.) The Springer Index of Viruses. Springer, Berlin.

90. Roberts, M. D., and H. Drexler. 1981. T1 mutants with increased transduction frequency are defective in host chromosome degradation. Virology 112:670–677.

91. Roberts, M. D., N. L. Martin, and A. M. Kropinski. 2004. The genome and proteome of coliphage T1. Virology 318:245–266.

92. Rodriguez-Casado, A., and G. J. J. Thomas. 2003. Structural roles of subunit cysteines in the folding and assembly of the DNA packaging machine (portal) of bacteriophage P22. Biochemistry 42:3437–3445.

93. Rohwer, F., and R. Edwards. 2002. The Phage Proteomic Tree: a genome-based taxonomy for phage. J. Bacteriol. 184:4529–4535.

94. Roncero, C., and M. J. Casadaban. 1992. Genetic analysis of the genes involved in synthesis of the lipopolysaccharide core in *Escherichia coli* K-12: three operons in the rfa locus. J. Bacteriol. 174:3250–3260.

95. Samsonov, V. V., V. V. Samsonov, and S. P. Sineoky. 2002. *DcrA* and *dcrB Escherichia coli* genes can control DNA injection by phages specific for BtuB and FhuA receptors. Res. Microbiol. 153:639–646.

96. Schanda-Mulfinger, U. E. and H. Schmieger. 1980. Growth of *Salmonella* bacteriophage P22 in *Escherichia coli dna*(Ts) mutants. J. Bacteriol. 143:1042–1045.

97. Schmieger, H., K. M. Taleghani, A. Meierl, and L. Weiss. 1990. A molecular analysis of terminase cuts in headful packaging of *Salmonella* phage P22. Mol. Gen. Genet. 221:199–202.

98. Schnaitman, C. A., and J. Klena. 1993. Genetics of lipopolysaccharide biosynthesis in enteric bacteria. Microbiol. Rev. 57:655–682.

99. Schneider-Scherzer, E., B. Auer, E. J. de Groot, and M. Schweiger. 1990. Primary structure of a DNA (N6-adenine)-methyltransferase from *Escherichia coli* virus T1: DNA sequence, genomic organization, and comparative analysis. J. Biol. Chem. 265:6086–6091.

100. Schoffler, H., and V. Braun. 1989. Transport across the outer membrane of *Escherichia coli* K12 via the FhuA receptor is regulated by the TonB protein of the cytoplasmic membrane. Mol. Gen. Genet. 217:378–383.

101. Sutton, M. D., K. M. Carr, M. Vicente, and J. M. Kaguni. 1998. *Escherichia coli* DnaA protein. The N-terminal domain and loading of DnaB helicase at the *E. coli* chromosomal origin. J. Biol. Chem. 273:34255–34262.

102. Tetart, F., C. Desplats, M. Kutateladze, C. Monod, H. W. Ackermann, and H. M. Krisch. 2001. Phylogeny of the major head and tail genes of the wide-ranging T4-type bacteriophages. J. Bacteriol. 183:358–366.

103. Tocchetti, A., G. Galimberti, G. Deho, and D. Ghisotti. 1999. Characterization of the oriI and oriII origins of replication in phage-plasmid P4. J. Virol. 73:7308–7316.

104. Traurig, M., and R. Misra. 1999. Identification of bacteriophage K20 binding regions of OmpF and lipopolysaccharide in *Escherichia coli* K-12. FEMS Microbiol. Lett. 181:101–108.

105. Vakharia, H., and R. Misra. 1996. A genetic approach for analysing surface-exposed regions of the OmpC protein of *Escherichia coli* K-12. Mol. Microbiol. 19:881–889.

106. Van de Peer, Y., and R. De Wachter. 1997. TREECON for Windows: a software package for the construction and drawing of evolutionary trees. Comput. Appl. Biosci. 10:569–570.

107. Vander, B. C., and A. M. Kropinski. 2000. Sequence of the genome of *Salmonella* bacteriophage P22. J. Bacteriol. 182:6472–6481.

108. Wagner, E. F., B. Auer, and M. Schweiger. 1983. *Escherichia coli* virus T1: genetic controls during virus infection, pp. 131–152. *In* M. Cooper, P. H. Hofschneider,

H. Koprowski, F. Melchers, R. Rott, H. G. Schweiger, P. K. Vogt, and R. Zinkernagel (eds.), Current Topics in Microbiology and Immunology. Springer, Berlin.

109. Wagner, E. F., H. Ponta, and M. Schweiger. 1977. Development of *E. coli* virus T1: repression of host gene expression. Eur. J. Biochem. 80:255–260.

110. Wagner, E. F., H. Ponta, and M. Schweiger. 1977. Development of *E. coli* virus T1: the pattern of gene expression. Mol. Gen. Genet. 150:21–28.

111. Watson, G., and K. Paigen. 1971. Isolation and characterization of an *Escherichia coli* bacteriophage requiring cell wall galactose. J. Virol. 8:669–674.

112. Werts, C., V. Michel, M. Hofnung, and A. Charbit. 1994. Adsorption of bacteriophage-lambda on the LamB protein of *Escherichia coli* K-22: point mutations in gene-*J* of lambda responsible for extended host-Range. J. Bacteriol. 176:941–947.

113. Whittam, T. S., S. D. Reid, and R. K. Selander. 1998. Mutators and long-term molecular evolution of pathogenic *Escherichia coli* O157:H7. Emerg. Infect. Dis. 4:615–617.

114. Wieczorek, D. J., and M. Feiss. 2001. Defining cosQ, the site required for termination of bacteriophage lambda DNA packaging. Genetics 158:495–506.

115. Williams, B. J., M. Golomb, T. Phillips, J. Brownlee, M. V. Olson, and A. L. Smith. 2002. Bacteriophage HP2 of *Haemophilus influenzae*. J. Bacteriol. 184:6893–6905.

116. Woods, D. E., J. A. Jeddeloh, D. L. Fritz, and D. DeShazer. 2002. *Burkholderia thailandensis* E125 harbors a temperate bacteriophage specific for *Burkholderia mallei*. J. Bacteriol. 184:4003–4017.

117. Wrobel, B., and G. Wegrzyn. 2002. Evolution of lambdoid replication modules. Virus Genes 24:163–171.

118. Wu, H., L. Sampson, R. Parr, and S. Casjens. 2002. The DNA site utilized by bacteriophage P22 for initiation of DNA packaging. Mol. Microbiol. 45:1631–1646.

119. Yethon, J. A., and C. Whitfield. 2001. Purification and characterization of WaaP from *Escherichia coli*, a lipopolysaccharide kinase essential for outer membrane stability. J. Biol. Chem. 276:5498–5504.

120. Young, R. 1992. Bacteriophage lysis: mechanism and regulation. Microbiol. Rev. 56:430–481.

121. Young, R., and U. Blasi. 1995. Holins: form and function in bacteriophage lysis. FEMS Microbiol. Rev. 17:191–205.

122. Zhang, N., and R. Young. 1999. Complementation and characterization of the nested Rz and Rz1 reading frames in the genome of bacteriophage lambda. Mol. Gen. Genet. 262:659–667.

123. Zuniga, M., B. Franke-Fayard, G. Venema, J. Kok, and A. Nauta. 2002. Characterization of the putative replisome organizer of the lactococcal bacteriophage r1t. J. Virol. 76:10234–10244.

18

T4 and Related Phages: Structure and Development

GISELA MOSIG
FRED EISERLING

History

Bacteriophages T2, T4, and T6 were among seven *Escherichia coli* phages selected by Max Delbrück to study fundamentals of viral replication in a limited number of model viruses. These studies led to the first formulation of many concepts that are now accepted foundations of molecular biology: the fundamental differences between growth of viruses and cells (figure 18-1) (109); the demonstration that nucleic acids of virus particles suffice to establish infection and to direct synthesis of complete virions (163); the concept of the gene, including distinctions between units of recombination, mutation, and function (30); genetic recombination as exchange between DNA molecules involving "heterozygous" overlaps (91, 164, 165); the demonstration of messenger RNA (mRNA) (55, 414) and the non-overlapping triplet code (83, 390) with nonsense triplets as termination signals (31); the repair of DNA damage in the light (103) and in the dark (149); restriction and modification of DNA (247); the presence of spliced and nonspliced introns in prokaryotes (25, 171); the definition of pathways leading to the assembly of complex macromolecular structures (105); and the importance of protein complexes (machines), which change composition during various DNA transactions (9, 275).

Many other phages from different parts of the world are classified within this family, based on similar sequences and map positions of their essential genes, regulatory patterns, virion structure and serological properties. They are now collectively called the "T-even" phages (192). More distantly related phages have been called pseudo-T-even phages (2, 13, 148, 266, 397) or schizo-T-even phages, although the criteria that distinguish them are ambiguous. These phages infect many different Gram-negative bacteria in various environments, from mammalian intestines to marine cyanobacteria and other bacteria.

Overview

The genomes of the T4-related phages are contained in large (~170,000 kb), linear, double-stranded (ds) DNA molecules whose termini contain repetitions of approximately 3% of the genome. The termini are randomly permuted over circular maps (figure 18-2) (273, 276, 289, 388, 389, 399). The DNA of most T4-related phages contains hydroxymethylcytosine (HMC) instead of cytosine. In most members of the family, the HMC residues are further glycosylated to different extents. These modifications together allow escape from host and phage restriction enzymes and are important for the developmental strategy of these phages.

DNA is packaged in elongated "heads" of quasi-icosahedral symmetry. DNA-filled heads are joined to independently assembled tails whose baseplates and fibers (figure 18-3) are instrumental for recognition, adsorption, and injection of the DNA into host bacteria. Different phages recognize different receptors in different host strains. Apparently, recombination within their genomes or with genomes of plasmids and other phages or prophages facilitates rapid evolution of T-even tail fiber genes and adaptation to different hosts (156, 396, 397).

T4 can grow well in many other Gram-negative bacteria if the bacteria are converted to spheroplasts and the phage are treated with urea (416). The urea circumvents the first adsorption stage (and host specificity) by altering the baseplates and tails of the phage particles so they can release their DNA into spheroplasts upon contact with bacterial inner membranes. The final injection of DNA requires membrane potential (135).

T-even phages are some of the most successful molecular parasites. Their genomes code for most phage-specific DNA replication, recombination, and repair functions. Because many of these proteins are similar in structure and function throughout all living organisms, the analysis of T4 proteins

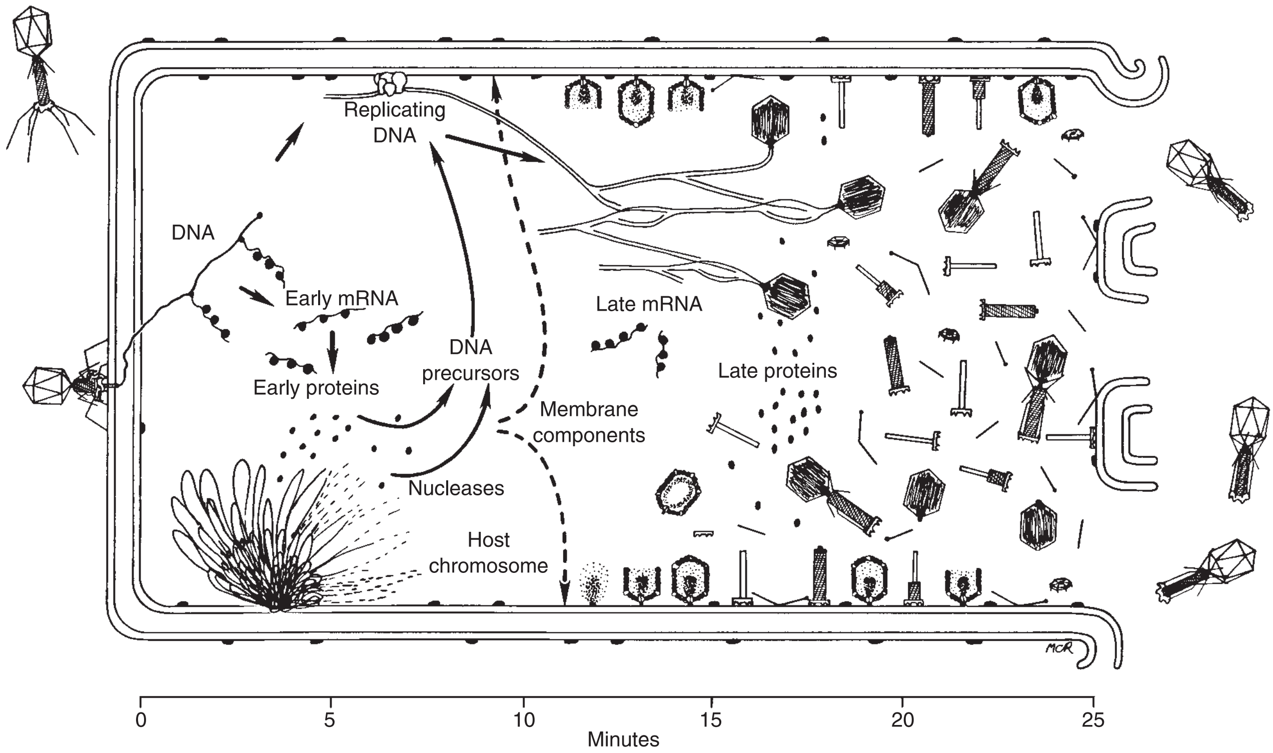


Figure 18-1 Overview of the T4 life cycle. Modified from (253).

has greatly contributed to general understanding of these processes.

Nevertheless, like all viruses, T-even phages depend for their propagation on many vital structures and functions of their hosts, such as membranes, energy metabolism, transcriptional and translational machines, and some chaperones. They manage to usurp host structures and functions in an exquisite choreography that allows adaptations to different environmental conditions, including the different physiological states of the host.

T4 is the most thoroughly investigated member of the T-even phages, mainly because the isolation of a large collection of conditional lethal mutants (110) provided a powerful impetus for molecular analyses by biochemical and biophysical methods. The results of the combined efforts of many research groups working on T4 and related phages are summarized in a monograph (192) and in a recent review that emphasizes bioinformatic aspects of the annotated T4 genome (262). Comparisons with other members of this phage family have been reviewed (87, 225, 226, 266) and will be extended as sequences of other T4-related phages are being published. References cited here are mainly to summarizing chapters in (192) and to papers published since then.

We emphasize that the recipe for T4's success as a molecular parasite is based on multiple redundant pathways for most, if not all, physiologically important DNA transactions: transcription, translation, replication, recombination, repair, and packaging. These pathways are interconnected at many levels (figure 18-4). Many of the cross-connections are promoted by certain proteins that can participate in different complexes and pathways. This allows communication between different pathways and their adaptation to different environmental conditions. The redundancies are most likely also important for evolution. However, for the sake of clarity, we discuss these different pathways and their regulation during phage development separately, after this overview.

T-even phages (in contrast to the T7- or N4-related phages (chapters 20 and 21) use the core host RNA polymerase throughout the infectious cycle, and the gradual subversion of host functions to support different aspects of phage propagation is achieved in many small steps:

1. A cascade of phage-induced proteins covalently and noncovalently modifies the host RNA polymerase and its accessory transcription factors (291). Together these modifications modulate processivity (termination) of elongating RNA polymerase and allow sequential recognition of different classes of promoters to selectively transcribe HMC-containing DNA. Thereby all host transcription is inactivated, and different classes of phage promoters are temporally controlled. At least one of the T4 proteins, an enzyme that ADP-ribosylates one α-subunit of host RNA polymerase, is packaged into virions and injected with the phage DNA into the next host bacterium. The sliding clamp required for DNA replication is also required for late

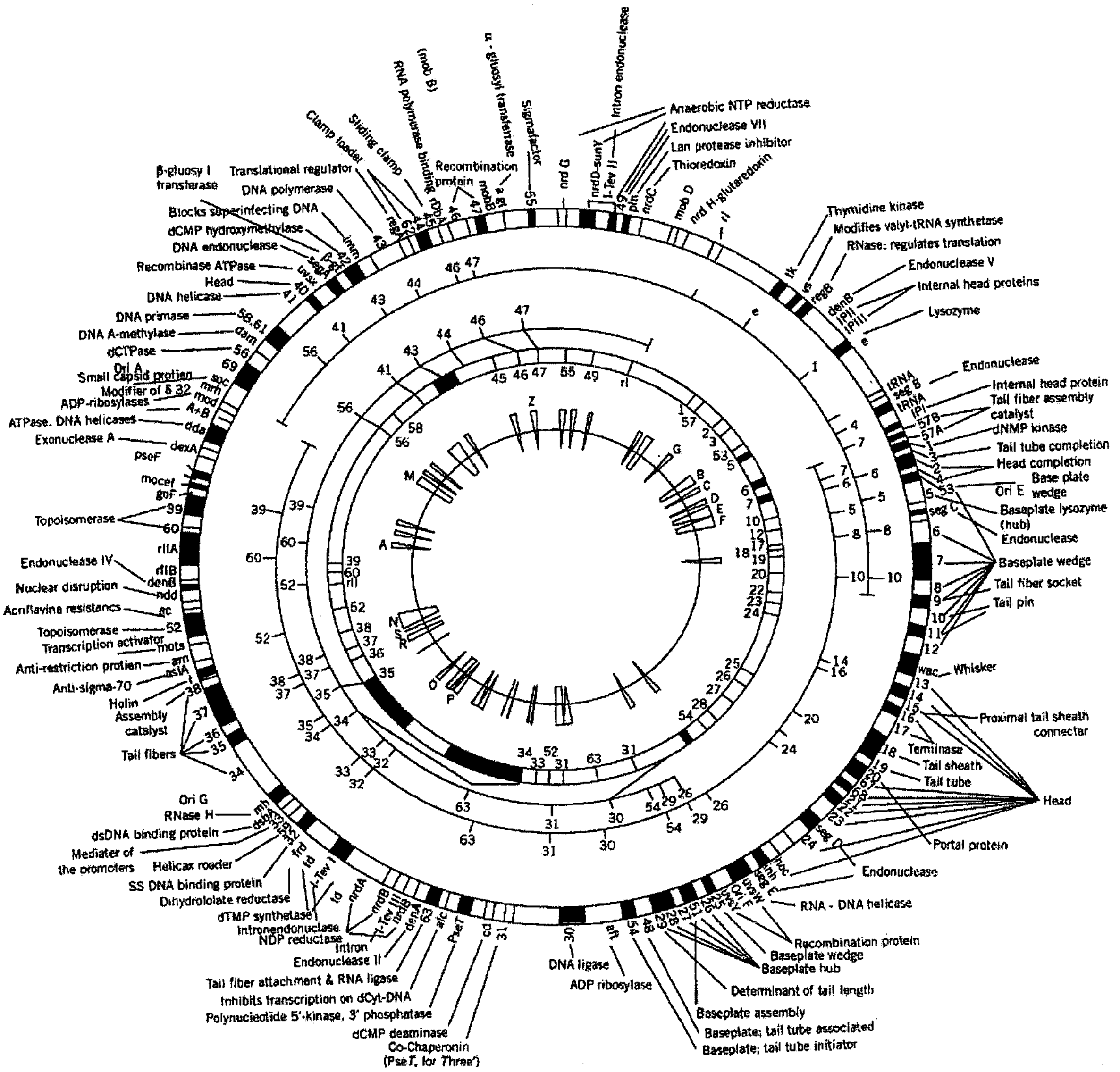

β-gluosy l transferase
Translational regulator
DNA polymerase
Clamp loader
Sliding clamp
RNA polymerase binding
(mob B)
Recombination protein
α - gluosyl transferase
Sigmafactor
nrd G
nrdD-sunY
I-Tev II
Intron endonuclease
Anaerobic NTP reductase
Endonuclease VII
Lan protease inhibitor
Thioredoxin
mob D
nrd H-glutaredoxin
Thymidine kinase
Modifies valyl-tRNA synthetase
RNase: regulates translation
Endonuclease V
Internal head proteins
Lysozyme
Endonuclease
Internal head protein
Tail fiber assembly catalyst
dNMP kinase
Tail tube completion
Head completion
Base plate wedge
Ori E
Baseplate lysozyme (hub)
Endonuclease
Baseplate wedge
Tail fiber socket
Tail pin
Whisker
Proximal tail sheath connectar
Terminase
Tail sheath
Tail tube
Head
Portal protein
Endonuclease
RNA - DNA helicase
Recombination protein
Baseplate wedge
Baseplate hub
Determinant of tail length
Baseplate assembly
Baseplate; tail tube associated
Baseplate; tail tube initiator
DNA ligase
ADP ribosylase
Co-Chaperonin (PseT, for Three')
dCMP deaminase
Polynucleotide 5'-kinase, 3' phosphatase
Inhibits transcription on dCyt-DNA
Tail fiber attachment & RNA ligase
Endonuclease II
Intron
NDP reductase
Intronendonuclase.
dTMP synthetase
Dihydrololate reductase
SS DNA binding protein
Helicax roader
Mediater of the promoters
dsDNA binding protein
RNase H
Ori G
Tail fibers
Assembly catalyst
Holin
Anti-sigma-70
Anti-restriction protien
Transcription activator
Topoisomerase
Acriflavine resistancs
Nuclear disruption
Endonuclease IV
Topoisomerase
Exonuclease A
ATPase. DNA helicases
ADP-ribosylases
Modifier of δ 32
Small capsid protien
Ori A
dCTPase
DNA A-methylase
DNA primase
DNA helicase
Head
Recombinase ATPase
DNA endonuclease
dCMP hydroxymethylase
Blocks superinfecting DNA
Z
G
B
C
D
E
F
M
A
N
S
R
O
P

Figure 18-2 T4 gene maps. The outer circle shows the approximate positions of characterized genes on the 169,903 base pair DNA, drawn as a circle. The next three genome segments show three maps derived from determining distances of chromosomal ends from rI, rII, and rIII respectively (278). Note that the relative length of the small rI molecules was 0.77, not 0.68 of the normal T4 chromosome. The next (full) circle indicates the positions of genes derived from genetic crosses (105). The innermost circle indicates the positions of heteroduplex loops after annealing of heat-denatured T2, T4, and T6 DNA (201).

transcription, connecting these two processes. The classification of transcripts is complex, as described below, largely because most genes are transcribed from multiple promoters, and protein synthesis is affected by subsequent translational controls.

2. RNA processing by phage and host enzymes, translational repressors, and poorly understood modifications of ribosomes all modulate T4 gene expression (69, 261). No transcriptional repressors (in the original meaning of the word) are known. Most likely, post-transcriptional modulators, including translational repressors, are more suitable than transcriptional regulators for adjusting to rapid physiological changes during the short T-even development: one growth cycle is completed in less than 30 minutes at 37 °C.
3. The host DNA and host mRNA, present at the time of infection, are rapidly degraded, and the breakdown products are efficiently reused to synthesize phage DNA and RNA.
4. The onset of the first phage DNA replication from specific origins requires host RNA polymerase to generate primers and is thereby influenced by the physiological regulatory processes of the host. Most

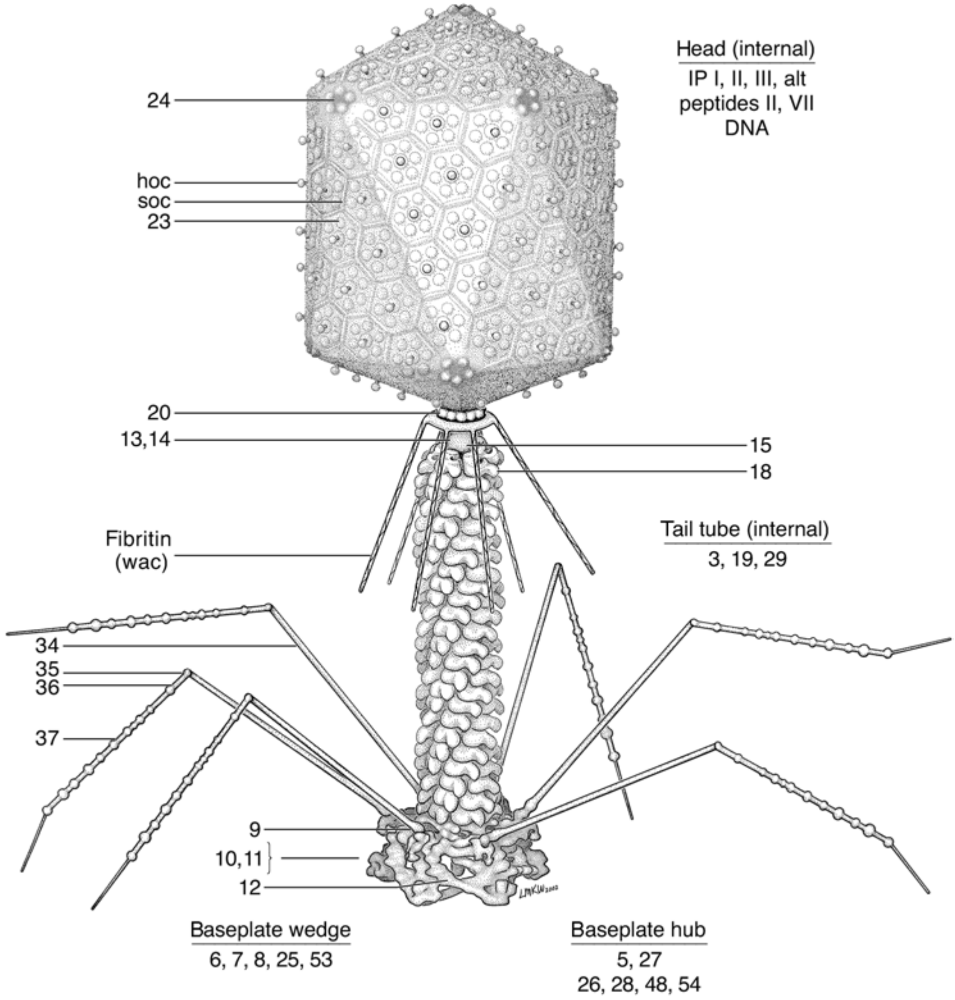


Figure 18-3 Structure of the T4 virion. Image is based on negative stain and cryo-electron microscopy, and crystallographic data. The locations of the protein components are indicated by gene number. The portal vertex composed of gp20 is attached to the upper ring of the neck structure, inside the head itself. The internal tail tube is inside the sheath and itself contains a structural component in its central channel. The baseplate contains short tail fibers made of gp12; these are shown in a stored or folded conformation.

subsequent T4 replication is entirely phage-controlled, escaping the host's controls. It depends on phage-encoded replication and recombination proteins and on primers that are intermediates of homologous recombination, thereby connecting replication, recombination, and repair of DNA (82, 217, 218, 277).

5. During the later stages of development, proteins involved in packaging DNA become also important for DNA replication, thereby coordinating these two processes (290).

Genome Structure and Map

The genome of T4 resides in 169,903 base pairs (bp) of dsDNA containing glucosylated HMC residues (262). The HMC residues are glycosylated to different extents in different T-even phages. The complex modification and restriction of T4 DNA, further discussed later in this chapter, can best be rationalized as the result of an ongoing evolutionary process that includes exchanges between the phage, its host, and prophages and plasmids resident in the host. Most T-even phages destroy dCTP, synthesize dHMCTP, and use the latter for their own DNA synthesis.

Using genetic tricks, T4 mutants containing unmodified cytosines in their DNA have been isolated (68, 228). Their DNA has been instrumental in cloning and sequencing the T4 genome, but the lack of modification affects restriction and regulation of transcription termination (98), and limits viability and host range (338). Moreover, as discussed below, origin initiation of DNA replication (439) and packaging (241) appear to be different from that of wild-type T4.

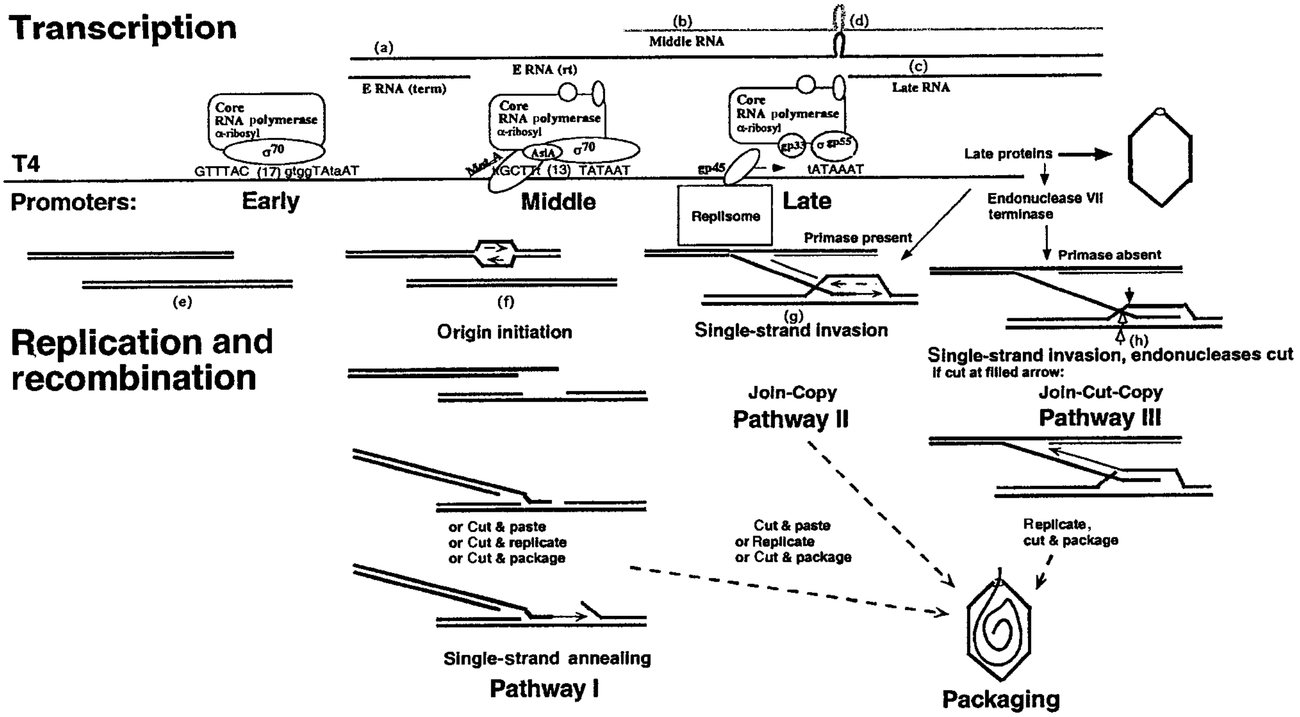

Transcription
(a)
(b)
Middle RNA
(d)
E RNA (rt)
E RNA (term)
(c)
Late RNA
Core
RNA polymerase
α-ribosyl
σ70
AsiA
gp33
σ gp55
gp45
T4
GTTTAC (17) gtggTAtaAT
TGCTTt (13) TATAAT
tATAAAT
Promoters:
Early
Middle
Late
Replisome
Late proteins
Endonuclease VII
terminase
Primase present
Primase absent
(e)
(f)
Origin initiation
(g)
Single-strand invasion
(h)
Single-strand invasion, endonucleases cut
If cut at filled arrow:
Replication and recombination
Join-Copy
Pathway II
Join-Cut-Copy
Pathway III
or Cut & paste
or Cut & replicate
or Cut & package
Cut & paste
or Replicate
or Cut & package
Replicate,
cut & package
Single-strand annealing
Pathway I
Packaging

Figure 18-4 Diagram of the relationship between the T4 transcriptional pattern and the different mechanisms of DNA replication and recombination. The upper panel shows the transcripts initiated from early, middle, and late promoters by sequentially modified host RNA polymerase. Hairpins in several early and middle transcripts inhibit translation of the late genes present on these messengerRNAs. The lower panel depicts the pathways of DNA replication and recombination detailed in this chapter. Hatched lines represent strands of homologous regions of DNA and the arrows point to possible positions of endonuclease cuts. Replication can be only initiated from cuts marked by filled arrows. Cuts indicated by open arrows cannot be used to initiate replication forks. Modified from (279).

Mature T4 DNA molecules (chromosomes) packaged into virions are linear. In contrast, intracellularly replicating DNA contains multiple covalently linked copies of the genome, called "concatemers" (117, 399), which are highly branched as a result of recombination and replication. These concatemers are cut during packaging (see below) at nearly random positions of the circular map (figure 18-2). Because the heads are filled with DNA corresponding to approximately 103% of theT4 genome, processive packaging generates the circular permutations of chromosomal ends (388) and ensures that each chromosome is diploid for more than 5000 bp at the ends (so-called terminal redundancies). The terminal redundancy allows homologous recombination between two terminal regions of the same chromosome and recombination-dependent DNA replication (85, 280) after infection of bacteria with a single phage particle.

Numerous mutations, and their assignments to complementation groups and open reading frames (ORFs), have defined approximately 160 proteins with known functions (table 18-1). More than 120 additional ORFs of unknown function can be deduced from the DNA sequence (262). Most, if not all of these ORFs are transcribed, and some of their protein products have been detected by polyacrylamide gel electrophoresis. There are many overlapping coding regions; several genes direct synthesis of more than one protein from in-frame or (in two cases) out-of-frame internal start codons; many ORFs are very small; and in at least one region (replication origin E) both complementary DNA strands encode proteins. In spite of the large genome size, only a few short regions are devoid of coding capacity. These and other observations suggest that apparently redundant and "nonessential" genes confer selective advantages.

In contrast to many other phages (e.g., the lambdoid and T7-related phages; chapters 27 and 20),T-even early and late transcription units are not clearly separated but are interdigitated (e.g., late genes *26*, *25* and the middle gene *uvsY*). Most genes are transcribed from multiple promoters, and genes for related or interacting proteins are not necessarily clustered (a most striking example is genes *5* and *27*, which code for the first interacting baseplate components; 187). Moreover, the direction of transcription does not unambiguously distinguish between early and late genes. In fact, in many cases early and late genes are cotranscribed (291) (figure 18-4). In this respect as well as in the amino acid sequences of some of their proteins (e.g., DNA polymerase and terminase), the T-even phages more closely resemble the herpesviruses than certain other phages.

Table 18-1 The Functions and Mutant Phenotypes of T4 Gene Products

Gene	Function of gene product	Size (kDa)	Mutant phenotype	Restrictive host or condition
rIIA	Membrane-associated protein; affect host membrane ATPase	82.9	Rapid lysis; suppress T4 *30* and some *32* mutations	*rex*$^+$ λ lysogens; P2-like HK239 lysogen*; tabR*
60	DNA topoisomerase subunit	18.6	DNA delay; *rc* = acriflavine resistance	25 °C or below
mobA	Pseudogene of Mob site-specific DNA endonuclease	4.2		Nonessential
39	DNA topoisomerase subunit; DNA-dependent ATPase; membrane- associated protein	58.0	DNA delay; *rc* = acriflavine resistance	25 °C or below; synthetic lethal with T4 *49* and *17* mutations, or when host topoisomerase IV is poisoned with novobiocin
*pla*CTr5x				CTr5x
goF = *com*C-α = *go9*H	Affects mRNA metabolism	16.7	Allows T4 growth in NusD *rho* hosts	Auxiliary
cef = *m b* = *M1*~*motC*	Processing of T4 tRNAs	8.5		Auxiliary; CT439; *roc*$^-$ hosts
*pse*F = plaCTr5x?	5′ phosphatase			Auxiliary
motB		18.2	Affects middle transcription	Auxiliary
dexA	Exonuclease A	26.0		Auxiliary; restricted on *optA*$^-$ hosts
dda = *sud*	DNA helicase; DNA-dependent ATPase	49.9	Suppress certain T4 *32* mutations	Auxiliary; synthetic lethal with T4 59 mutations
srd = *dda.2*	Postulated decoy of host σ^{70} or σS	29.0		Auxiliary
modA	Adenylribosylating enzyme	23.4	α-subunits of host RNA polymerase are incompletely modified	Auxiliary
modB	Adenylribosylating enzyme	24.2		Auxiliary
srh = *modA*.5	Postulated decoy of host σ^{32}	8.1	Delays early T4 gene expression at high temperatures	Auxiliary
mrh	Affects phosphorylation of host σ^{32}	18.5	Allows T4 growth in a σ^{32} host	Auxiliary
soc	Small outer capsid protein	9.1	Unstable T4 capsids	Auxiliary
segF = *69*	Intron-like endonuclease. A probable fusion protein, generated from *56* and *69* by hopping of ribosomes across a pseudoknot, is larger	26.2		Nonessential
56	dCTPase; dUTPase; dCDPase; dUDPase	20.4	Little DNA synthesis; unstable DNA	Essential
oriA	several sequences in *56*, *69*, and *soc* required in *cis*; primer transcript same as transcripts for these genes		No replication from origin A	Auxiliary
dam	DNA adenine methylase	30.4	No DNA adenine methylation	Auxiliary
61 = *58*	Primase; requires interaction with gp41 helicase for priming at unique sequence	39.8	DNA delay	Auxiliary; 25 °C or below; synthetically lethal with T4 *49* or *17* mutations
sp = *61.3*~ *rIV*	Periplasmic protein	11.0	Rapid lysis; suppresses *e* lysozyme mutations	Auxiliary
dmd = *61.5*	Discriminator of mRNA degradation	7.0	Excessive mRNA degradation	Nonessential; suppressed by *motA* mutations
41	Replicative and recombination DNA helicase; GTPase; ATPase; dGTPase; dATPase	53.6	DNA arrest; little DNA displacement synthesis	Essential
40	Membrane-associated protein initiator of head vertex	13.3	Polyheads	Auxiliary; high temperatures
uvsX = *fdsA*	RecA-like recombination protein; DNA-ATPase	44.0	UV- and X-ray sensitive; recombination deficient; suppress *49* mutations	Auxiliary
segA	Site-specific intron-like DNA endonuclease	25.3		Nonessential
β-gt	β-glucosyltransferase	40.7	No β-glucosylation of HMC DNA	Auxiliary; *Shigella*

(Continued)

Table 18-1 Continued

Gene	Function of gene product	Size (kDa)	Mutant phenotype	Restrictive host or condition
42	dCMP hydroxymethylase	28.5	Little or no DNA synthesis	Essential
imm	Inner membrane protein	9.3	No immunity to superinfection, membrane protein	Auxiliary
43	DNA polymerase; 3′ to 5′ exonuclease	103.6	No DNA synthesis; mutator or antimutator activities of conditional lethals under semipermissive conditions	Essential; nonessential *dsd* mutants do not grow in *optA* hosts
regA	Translational repressor of several early genes	14.6	Extended synthesis of several early proteins	Auxiliary; restricted in *rpoB5081* at 42 °C
62	Clamp-loader subunit	21.4	No DNA synthesis	Essential
44	Clamp-loader subunit	35.8	No DNA synthesis	Essential
45	Processivity enhancing sliding clamp of DNA polymerase; and mobile enhancer of late promoters	24.9	No DNA synthesis; no late transcription	Essential
rpbA	RNA polymerase binding protein	14.7		Auxiliary
46	Recombination protein and nuclease subunit	63.6	Recombination deficient; DNA arrest; no host DNA degradation	Essential in B strains; mutants are ''leaky'' in some K strains
47	Recombination protein and nuclease subunit	39.2	Recombination deficient; DNA arrest; no host DNA degradation	Essential in B strains; mutants are ''leaky'' in some K strains
α-gt	α-glucosyl-transferase	46.7	No α-glucosylation of HMC	Auxiliary
mobB	Putative site-specific intron-like DNA endonuclease	30.4		Nonessential
55	σ factor recognizing late T4 promoters	21.5	No late transcription	Essential
nrdH = *55.7*	Anaerobic nucleotide reductase subunit	11.7		Auxiliary
nrdG = *55.9*	Anaerobic nucleotide reductase subunit	18.2		Auxiliary
mobC = *55.10*	Putative site-specific intron-like DNA endonucleasee	24.0		Auxiliary
nrdD = *sunY*	Anaerobic ribonucleotide reductase subunit; RNA contains a self-splicing intron	68.064		Anaerobic growth
I-Tev II	Endonuclease for *nrdD*-intron homing	30.4		Auxiliary
49	Recombination endonuclease VII	18.1	No resolution of recombination junctions; incomplete packaging of DNA; reduced heteroduplex repair, reduced DNA synthesis	Essential
49′	Internal translation initiation product	11.9		
pin	Inhibitor of host Lon protease	18.8	No degradation of amber peptides	Auxiliary
nrdC	Thioredoxin, glutaredoxin	10.1		Auxiliary
mobD	Putative site-specific DNA endonucleasee	30.5		Nonessential
rI = *tk.-2*	Membrane protein	11.1	No lysis inhibition	Auxiliary
tk	Thymidine kinase	21.6		Auxiliary
vs	Modifier of valyl-tRNA synthetase	13.1		Auxiliary
regB	Site-specific RNase	18.0	Misregulation of early genes	Auxiliary
denV	Endonuclease V; *N*-glycosidase	16.1	UV-sensitive	Auxiliary
ipII	Internal protein II	11.1--9.9		Auxiliary
ipIII	Internal protein III	21.7--20.4		Auxiliary
e	Soluble lysozyme; endolysin	18.7	No cell lysis	Essential, except when suppressed by *sp* and *5* mutations
nudE = *e.1*	Nudix hydrolase	17.0		Auxiliary
goF3			Allow T4 growth in *nusD rho* hosts	Auxiliary
rnaC = Species 1	Stable RNA			Auxiliary
rnaD = Species 2	Stable RNA			Auxiliary
$tRNA^{arg}$			psu_4 opal suppressor	CT439
segB	Probable site-specific intron-like DNA endonuclease	26.2		Nonessential

(Continued)

Table 18-1 Continued

Gene	Function of gene product	Size (kDa)	Mutant phenotype	Restrictive host or condition
tRNA^{ile}				CT439
tRNA^{thr}				CT439
tRNA^{ser}			psu$_a$; psu$_b$; psu$_1$; amber suppressors	CT439
tRNA^{pro}				CT439
tRNA^{gly}				CT439
tRNA^{leu}			psu$_3$	CT439
tRNA^{gln}			psu$_2$; SB	CT439
ipI	Internal protein I	10.2--8.5		CT596
57B		17.2		?
57A	Chaperone of long and short tail fiber assembly	8.7	Defective tail fiber assembly	Essential; by-passed by certain host mutations
1	dNMP kinase	27.3	No DNA synthesis	Essential
3	Head-proximal tip of tail tube	19.7	Unstable tails	Essential
2 = 64	Protein protecting DNA ends	31.6	Noninfectious particles with filled heads	Essential, except in *recBCD* hosts
4 = 50 = 65	Head completion protein	17.6	Noninfectious particles with filled heads but tails attached at wrong angles	Essential
53	Baseplate wedge component	23.0	Defective tails	Essential
5	Baseplate lysozyme; hub component	63.1--44 &19	Defective tails	Essential
oriE	*cis*-acting sequences in genes *4*, *53*, *5*; primer transcript in opposite orientation of gene *5* transcripts		No DNA replication from *oriE*	Auxiliary
repEB	Protein required for initiation from *oriE*	5.48	No DNA replication from *oriE*	Auxiliary; synthetic lethal with *motA* mutation
repEA	Protein auxiliary for initiation from *oriE*	6.13	Anomalous DNA replication from *oriE*	Auxiliary
segC	Site-specific intron-like DNA endonuclease	22.2		Nonessential
6	Baseplate wedge component	74.4	Defective tails; permit plating of fiberless phage	Essential
7	Baseplate wedge component	119.2	Defective tails; permit plating of fiberless phage	Essential
8	Baseplate wedge component	38.0	Defective tails	Essential
9	Baseplate wedge component, tail fiber socket, trigger for tail sheath contraction	31.0	No attachment of tail fibers	Essential
10	Baseplate wedge component, tail pin	66.2	Defective tails	Essential
11	Baseplate wedge component, tail pin, interface with short tail fibers, gp12	23.7	Defective tails	Essential
12	Short tail fibers	56.2	Defective tails	Essential
wac	Whiskers, facilitate long tail fiber attachment	51.9	No whiskers	Auxiliary
13	Head completion	34.7	Inactive, but filled heads	Essential
14	Head completion	29.6	Inactive, but filled heads	Essential
15	Proximal tail sheath stabilizer, connector to gp3 and/or gp19	31.6	Defective tails	Essential
16	Terminase subunit, binds double-stranded DNA;	18.4	Empty heads	Nearly essential
16′	Truncated C-terminal end			
17	Terminase subunit with nuclease and ATPase activity; binds ssDNA, gp16 and gp20	69.8	Empty heads	Essential
17′A	Terminase subunits with nuclease and ATPase activity; internal transcription and translation in frame; does not bind ssDNA	59.2		
17′B		57.1		

(Continued)

Table 18-1 Continued

Gene	Function of gene product	Size (kDa)	Mutant phenotype	Restrictive host or condition
17″	Terminase subunit with nuclease and ATPase activity (transcript processing and internal initiation of translation in frame); does not bind ssDNA. Several additional proteins most likely initiated from internal ribosome binding sites of the *17* transcripts	46.8		
18	Tail sheath monomer	71.3	Defective tails	Essential
19	Tail tube monomer	18.5	Defective tails	Essential
20	Portal vertex protein of the head	61.0	Polyheads	Essential
pip = *67*	Prohead core protein; precursor to internal peptides	9.1-- small peptides	Defective heads	Essential
68	Prohead core protein	15.9	Isometric heads	Essential
21	Prohead core protein and protease	23.3-- small peptides,	No or defective heads	Essential
21′	Prohead core protein and protease (internal initiation of translation)	20.8-- small peptides	Defective heads	
22	Prohead core protein; precursor to internal peptides	29.9-- small peptides	No or faulty heads	Essential
23	Precursor of major head subunit	56.0--48.7--43	No or faulty heads; *gol* mutations in gene *23* allow growth in *lit* hosts (CTR5x)	Essential; Gol peptide together with *E. coli* Lit, cleaves host EF Tu
segD	Probable site-specific intron-like DNA endonuclease	25.6		Nonessential
24 = *os*	Precursor of head vertex subunit	47.0--46, 48.4?	No or faulty heads, osmotic shock resistance	Essential; by-passed by certain gene *23* mutations
rnlB = *24.1*	Second RNA ligase	37.6		?
hoc = *eph*	Large outer capsid protein	40.4	Unstable capsids	Auxiliary
inh = lip	Minor capsid protein; inhibitor of gp21 protease	25.6		Auxiliary
segE	Probable site-specific intron-like DNA endonuclease	22.9		Nonessential
uvsW = *dar*	RNA–DNA- and DNA-helicase; DNA-dependent ATPase	67.5	UV-sensitive; fail to unwind R-loops; suppress T4 *59 uvsX*, *uvsY*, and *46* mutations	Auxiliary
uvsY = *fdsB*	ssDNA binding, recombination and repair protein; helper of UvsX, inhibitor of endo VII	15.8	UV-sensitive; recombination-deficient; repair-deficient, DNA arrest; suppress T4 *49* mutations	Auxiliary
oriF = *oriuvsY*	*cis* acting sequences in genes *uvsY*, *uvsY.-1* and *uvsY.-2*; primer transcript same as *uvsY*, *uvsY.-1* and *uvsY.-2* transcript		No DNA replication from *oriF*	Auxiliary
25	Baseplate wedge subunit	15.1	Defective tails	Essential
26	Baseplate hub subunit	23.9	Defective tails	Essential
26′	Internal in-frame translation initiation	12		?
26″	Internal out-of-frame translation initiation	10		?
51	Baseplate hub assembly catalyst?	29.3	Defective tails	Essential
27	Baseplate hub subunit	44.5	Defective tails; permit plating of fiberless phage	Essential
28	Baseplate distal hub subunit	17.3	Defective tails	Essential
29	Baseplate hub; determinant of tail length	64.4	Defective tails	Essential
48	Baseplate; tail tube associated	39.7	Defective tails	Essential
54	Baseplate-tail tube initiator	35.0	Defective tails	Essential

(Continued)

Table 18-1 Continued

Gene	Function of gene product	Size (kDa)	Mutant phenotype	Restrictive host or condition
alt	Adenosylribosyltransferase (packaged and injected with DNA)	75.8	Synthetic defective with *modA* and *modB* deletions	Auxiliary
30 = lig	DNA ligase	55.3	DNA arrest; hyper-recombination	Essential; can be by-passed by functioning host ligase, when T4 *rII* is defective
rIII	unknown	9.3	Rapid lysis	Auxiliary
31	Co-chaperonin for GroEL	12.1	Head assembly; gp23 forms lumps; T4 topoisomerase is defective	Essential
cd	dCMP deaminase	21.2		Auxiliary
pseT	Deoxyribonucleotide 3′ phosphatase, 5′ polynucleotide kinase	34.6		Auxiliary; CTr5x (*lit*$^{-}$)
alc = unf	RNA polymerase- and DNA-binding protein; transcription terminator on dC-DNA	19.0	Allow transcript elongation on C-DNA; no unfolding of host nucleoid	*E. coli* (pR386)
rnlA = 63	RNA ligase; catalyst of tail fiber attachment	43.5	Defective tail fiber attachment	Auxiliary
denA	Endonuclease II that restricts dC-containing DNA	16.7	Defective in host DNA degradation	Auxiliary; restricted in *E. coli* B *rpoB5081*
nrdB	Ribonucleotide reductase β subunit (contains intron)	45.3	Reduced DNA synthesis	Auxiliary; *nrd*-defective hosts
I-TevIII	Defective intron homing endonuclease	11.3		Nonessential
mobE	Putative mobile endonuclease	16.5		Nonessential
nrdA	Ribonucleotide reductase α-subunit	86	Reduced DNA synthesis	Auxiliary; *nrd*-defective hosts
td	Thymidylate synthetase (contains intron)	33.1	Reduced DNA synthesis	Auxiliary; *td*-defective hosts
I-TevI	Intron homing endonuclease	28.2		Auxiliary
frd	Dihydrofolate reductase	21.7	Reduced DNA synthesis	Auxiliary
32	ssDNA binding protein, scaffold of DNA replication, recombination and DNA-precursor-synthesizing protein machines	33.5	DNA arrest, UV-sensitive, recombination and excision repair deficient	Essential; Tab32for *ts* mutants; *32 am* mutations in ochre-suppressor-containing hosts are suppressed by *dda* mutations
segG = 32.1	Site-specific DNA endonuclease, leading to localized gene conversion, exclusion	24.6		Auxiliary
59	Loader of gene *41* DNA helicase, ssDNA binding protein	26.0	Fail to load gp41 helicase onto recombination intermediates, or ssDNA covered with gp32 or UvsX protein; DNA arrest	Almost essential
33	Protein connecting gp45 and gp55, to allow transcription by RNA polymerase from late promoters	12.8	No late RNA synthesis	Essential
dsbA	Double-stranded DNA binding protein	10.4	Facilitates some late RNA synthesis	Auxiliary
rnh = das	RNaseH; 5′ to 3′ DNase; yeast FEN homolog	35.6	Defective processing of Okazaki fragments; *das* mutations suppress T4 *46*, *47* and *uvsX* mutations	Auxiliary
34	Proximal tail fiber subunit	140.4	Fiberless particles	Essential
oriG = ori34	Primer transcript in opposite orientation of *34* transcript			Auxiliary
35	Tail fiber hinge	40.1	Fiberless particles	Essential
36	Small distal tail fiber subunit	23.3	Fiberless particles	Essential
37	Large distal tail fiber subunit	109.2	Fiberless particles, host range	Essential
38	Assembly catalyst of distal tail fiber	22.3	Fiberless particles	Essential
t = rV∼stII	Holin, inner membrane pore protein, affects lysis timing and inhibition	25.2	Affect lysis by *e* lysozyme; suppress T4 *rII* and *63* mutations	Essential

(Continued)

Table 18-1 Continued

Gene	Function of gene product	Size (kDa)	Mutant phenotype	Restrictive host or condition
asiA	Protein that binds to host σ^{70}, inhibits interaction with −35 regions of classical promoters, and facilitates interaction with T4 MotA protein	10.6	Defective middle mode, and (indirectly) late transcription	Almost essential
arn	Inhibitor of MrcBC restriction nuclease	10.9		Auxiliary
motA = *sip*	Activator of middle promoters; dsDNA binding protein specific for mot boxes	23.6	Defective middle mode transcription; suppress *rII*-defects in λ lysogens; affects interaction with σ^{70} and AsiA	Almost essential
52	DNA topoisomerase subunit; membrane-associated protein	50.6	DNA delay	Temperatures below 25 °C; inhibition of host topoisomerase IV with novobiocin
ac	Membrane protein	5.5	Acriflavine resistant	Auxiliary
ama∼ *rs*		5.4	Acriflavine resistant	Auxiliary
stp	Peptide modulating host restriction system	3.7	Suppress *pseT* mutations	Auxiliary
ndd = D2b	Protein that disrupts host nucleoid; binds to host HU	16.9	Nucleoid disruption defective	Auxiliary; CT447
pla262	Unknown			CT262
denB	Endonuclease IV, single-strand specific endonuclease	21.2	Allows progeny production of T4 with dC-DNA	Auxiliary
rIIB	Membrane-associated protein; affects host membrane ATPase	35.5	Rapid lysis; suppresses T4 *30* and some 32 mutations	*rex*$^+$ λ lysogens; P2-like HK239 lysogen; *tabR*

Genes are listed by the currently used names, followed by alternative designations in the literature. Gene products processed into smaller peptides are indicated (- -) with the sizes or size range following the principal product. Because the distinction between "essential" and "nonessential" is not obvious, when the mutants were not tested under all possible conditions or in all possible hosts, "nonessential" genes are noted as auxiliary. Where known, restrictive hosts or plating conditions for several mutant genes are noted.

What is the meaning of these complexities and redundancies? It is useful to think about them in terms of phage evolution (153) (chapter 4). The T4 genome is assembled by mixing and matching different components, entire genes and gene segments, from the genomes of other phages, plasmids, or hosts, and evolutionary pressure is still at work to coordinate these components into a functional unit.

Virion Structure and Initiation of Infection

Phage T4 devotes a large percentage of its genes to the synthesis and assembly of structural components, from which the complex virion is assembled. Figure 18-3 shows a diagram of the T4 particle derived from negative stain, cryo-electron microscopy and X-ray diffraction, and lists the structural components of the phage. Twenty-four genes control head morphogenesis. The head itself contains at least 10 different polypeptides (figure 18-3), while other T4 genes control prohead initiation (gp20, gp40), protein folding (gp31), scaffold assembly (gp22, gp67, gp68, lip = inh), and proteolytic processing (gp21), DNA packaging and protection (gp16, gp17; gp2) and head completion (gp13, gp14, gp4). Gp wac is fibritin, from which the whiskers are assembled.

The tail and fibers contain 26 structural proteins, and five other T4 gene products are needed as assembly catalysts or chaperones: gp51, gp63, gp57A, gp31, and gp38.

The long tail fibers, attached to the baseplate, and fibritin (wac) "whiskers" near the head serve as sensors to detect the presence of a bacterial cell. The baseplate–tail complex is a molecular machine designed to deliver DNA into the cell. Like a cocked mechanical toy, it stores energy in the form of constrained protein conformations.

T4 absorption and infection occur in several steps. Before the T4 infectious cycle begins, the long tail fibers are released from their stored position at the whiskers. This permits the thin tips of the long fibers to interact reversibly with the cell surface, at diglucosyl residues in the outer membrane in *E. coli* B lipopolysaccharide (LPS) (135) or at the OmpC protein in *E. coli* K-12. Phage T4 uses gp37 to attach to the host's cell surface, while phage T2 and several other T-even phages use gp38. In phage T2, fiber attachment is to the bacterial surface protein OmpF, while phage K3 uses the OmpA protein (156, 436).

The long tail fibers, while binding reversibly to the outer membrane receptors, move over the cell surface and transmit conformational information to the baseplate via gp9, which normally keeps the baseplate in a constrained position (84). Binding by at least three long tail fibers converts the baseplate from a hexagon into an expanded, star-like structure that releases the gp12 short tail fibers from their stored and folded position. This leads to irreversible binding of the virus to the host cell by contact of the gp12 short tail fibers with a surface receptor, using heptose residues in the *E. coli* LPS core. The short tail fibers of all T-even phages bind to surface LPS but not to the outer membrane proteins. Sheath contraction is also regulated in part by the long tail fibers, since fiberless phages are more resistant to chemically induced sheath contraction than are phages with the normal complement of long tail fibers.

Several models have been described for the nature of the protein interactions during sheath contraction (107). The "induced conformational change" model of (195) proposes that sheath extension is maintained by sheath–tube subunit interactions, and that there is no tension along the sheath. Contraction is triggered by a wave of conformational change along the tube that releases the gp18–gp19 interactions, and sheath contraction follows this wave. This model incorporates Moody and Markowski's (268) observations of partially contracted sheaths and the artificial crosslinking of sheath to tube by formaldehyde treatment. A combination of models, with conformational energy supplied by gp18–gp19 interactions and initiation by release of gp18–gp19 interactions, has been presented by Caspar (70). This model shows in detail the steps in the contraction of a mechanical device that illustrates clearly the structural relationships during the contraction process.

A needlelike structure centered under the baseplate, composed of gp27 and gp5, has been proposed to initiate penetration of the cell envelope (187). Gp5, which is cleaved during this process, contains a lysozyme domain that may locally digest the cellular peptidoglycan, preparing the way for DNA injection (292). The needle must then move aside or detach to allow penetration of the cell wall by the tail tube. The force required for penetration is associated with contraction of the tail sheath.

The last step in infection, DNA transfer, remains poorly characterized, but most likely involves a transmembrane channel, evidenced by ion leakage from host cells upon infection (135). A putative channel may pre-exist in the membrane, or form from host proteins, phage proteins, or a combination of both. Its function requires membrane potential (135).

Alternatively, it has been proposed that the tail tube itself acts as a DNA conduit directly into the host cell. Studies by Furukawa et al. (124, 125) and Tarahovsky et al. (395) suggest that the tail tube penetrates the outer membrane by mechanical force generated by sheath contraction, then induces localized inner membrane fusion around the tip of the tail tube.

Several virion proteins are ejected with the DNA. The gp29 "tape measure" protein, located in the tail tube channel, must exit before DNA can enter. Gp2 protects the ends of the phage DNA from exonuclease V (370). The ADP ribosylating protein, gp alt, may influence the earliest stage of T4 transcription (206). The cleaved internal proteins are also injected, but their intracellular roles are unknown (50).

Complete transfer of phage DNA and proteins into a host cell requires energy supplied by the membrane-potential component of the proton-motive force (135). However, membrane potential is not the main driving force of DNA transfer, since artificially contracted phage can transfer their DNA into de-energized spheroplasts (124). It is proposed that the role of membrane potential is to provide energy to bring the outer and inner membranes closer together to facilitate interaction of the phage tail with the inner membrane, and that energy regulates, but does not drive DNA transfer.

Temporally Controlled Transcription

The Hershey–Chase experiment established unambiguously that T-even DNA is sufficient to establish viral growth (163). Inactivation of host functions and temporal regulation of expression of different classes of phage genes are then exerted at several levels: transcript initiation, elongation, termination, and stability; translational controls, and combinations thereof (227, 261, 291).

In contrast to the T7- or N4-related phages, the T-evens use the host's core RNA polymerase throughout their life cycle. Differential transcription initiation from T4 early, middle or late promoters is due, in part, to a cascade of RNA polymerase modifications: several T4 proteins (table 18-1) associate noncovalently with the core RNA polymerase, and the α-subunits are covalently ADP-ribosylated. Several of these proteins appear "nonessential" under standard laboratory conditions, suggesting that the transitions from host gene expression to different stages of T4 development occur in many small and redundant steps (288, 209).

One of the earliest synthesized T4 proteins, Alc, binds to RNA polymerase and DNA, and selectively terminates transcription from C-containing (host), but not HMC-containing (T4) DNA (98, 194, 376). This causes unfolding of host DNA, explaining the alternative name *unf* for the corresponding gene (161). The inactivation of host transcription is subsequently reinforced by other modifications of the host's RNA polymerase.

The developmental regulation of T4 gene expression more closely resembles a web of interacting regulatory networks than a cascade of linear timed pathways. This has led to often confusing nomenclature. Different classes of phage genes can be distinguished as early (also called immediate early, IE), middle (also called delayed early, DE), or late. Genes that are transcribed both early and late have also been called quasi-late.

Another classification criterion distinguishes all genes that are expressed prior to the onset of DNA replication ("prereplicative" or "early") from "postreplicative" or "late" genes, whose expression depends on DNA replication, specifically the sliding clamp gp45, on a T4-encoded sigma factor (gp55) and on a mediator protein, gp33.

Early Transcription

Operationally, early (IE) genes are defined as being transcribed by host RNA polymerase when phage gene expression is prevented by inhibitors of protein synthesis (e.g., chloramphenicol). Under these conditions the host RNA polymerase uses unmodified σ^{70} to recognize early T4 promoters, but the core RNA polymerase may be altered by T4 Alt protein that is injected with the phage DNA. Alt protein ADP-ribosylates one α-subunit at Arg265. This amino acid is important for recognition of so-called UP elements in promoters for ribosomal RNA (343), for several activator proteins of the host and of other phages (174), and for the dimerization of the C-terminal domains of RNA polymerase α-subunits (62, 126). Such ADP-ribosylation is beneficial but not essential for T4 growth under optimal laboratory conditions.

Most T4 early promoters resemble the consensus sequence of the major *E. coli* promoters, but they have extended −10 regions, different −35 regions, and additional information content (430, 431) (figure 18-4). Many of them have upstream poly (A) or poly (T) tracts, which may enhance transcription as bendable sequences or as UP elements or both. At least three exceptional early T4 promoters (P_{57}, P_{bac}, and P_{repE}) more closely resemble the major *E. coli* promoters (162, 239, 407).

Several factors, none of them essential for T4 growth, have been proposed to contribute to the preferential transcription from T4's early promoters when *E. coli* promoters are not yet degraded (227).

1. Most importantly, the early T4 Alc protein, once it is made, inhibits all transcript elongation on (cytosine-containing) host DNA.
2. ADP-ribosylation of Arg^{265} of one α-subunit of the host's RNA polymerase by the T4 Alt protein and by the early T4 ModA protein differentially affects transcription from various T4 or *E. coli* promoters (430, 431).
3. At least two small T4 proteins (Srh and Srd) share interaction sites of two *E. coli* sigma factors with the core RNA polymerase. They might compete with them like decoys (288).
4. At the time of infection, the host DNA is associated with nonspecific (e.g., HU, H-NS) or semispecific (e.g., IHF, FIS) DNA binding proteins. In contrast, the infecting phage DNA is at first largely free of proteins and may be more readily accessible to the host's RNA polymerase (227).
5. The early T4 AsiA protein binds to the C-terminal segment of σ^{70} (166, 230, 314, 356), interfering with transcription from host and phage promoters containing consensus −35 regions (5, 6, 77, 209, 313).

Later on, the degradation of host DNA by several early T4 proteins (66, 229, 375) eliminates all host transcription.

Middle Transcription

In contrast to early (IE) genes, expression of middle (and late) T4 genes requires phage-directed protein synthesis for several reasons. Transcription from middle promoters requires prior synthesis of two early T4 proteins: AsiA and MotA. Alternatively, downstream genes of transcription units initiated from early promoters may require antitermination or stabilization of rare long transcripts (383–385). The distinction between different classes is blurred, however, because many T4 genes are under dual or multiple transcriptional controls, and transcription of middle (DE) genes may be delayed for several reasons. Moreover, many late T4 genes are transcribed from late promoters but are also cotranscribed with early genes from early or middle promoters. For these late genes, post-transcriptional mechanisms (discussed below) prevent expression at early times.

The AsiA protein (77, 166–168, 382) inhibits host transcription, because it binds to region 4 of σ^{70} (265, 316, 355, 356) and prevents transcription of host promoters with "standard" −35 regions. It has been postulated that AsiA also turns off early T4 promoters. However, experiments using *asiA* deletions indicate that other, yet unknown factors are important for the shutoff of most early T4 promoters (325), which have nonstandard −35 regions.

AsiA is also an activator of T4 middle promoters (166). At middle promoters the σ^{70} subunit, complexed with AsiA protein and core RNA polymerase, recognizes consensus −10 regions, and the AsiA protein interacts with T4 MotA protein (255) bound to consensus DNA sequences centered at −30 positions, called *motA* boxes (166, 251, 383) (figure 18-4). AsiA is a dimer in solution, but a monomer when associated with σ^{70}. The dimer interface of one monomer is alternatively used to contact σ^{70} (230). Two sequenced T4-related phages, RB49 (87) and KVP 40 (262), apparently lack the middle mode transcription.

Collectively, prereplicative genes encode: (i) nucleases that degrade the host DNA; (ii) enzymes of the deoxyribonucleotide biosynthesis complex; (iii) proteins of the replication and recombination machines; (iv) proteins that modify the T4 DNA to protect it from degradation by its own nucleases and from other restriction enzymes; (v) several tRNAs, which are processed from precursor RNAs and supplement host tRNAs during translation; (vi) proteins that modify structure and function of the host RNA polymerase; (vii) at least two RNases: RegB protein that selectively destroys certain early and probably host transcripts, and RnaseH, that degrades RNA primers (307); (viii) a differential modulator of late RNA degradation (Dmd) (184–186); and (ix) a translational repressor (RegA protein) (261, 429). In addition, some prereplicative transcripts serve as primers for leading-strand DNA synthesis in origins of replication (see below).

Late Transcription

Transcription from late promoters requires concomitant DNA replication and the replacement of σ^{70} with a phage-encoded sigma factor, gp55, the product of a middle gene (432). Initiation from late promoters (129) (figure 18-4) also requires the adapter protein, gp33, and the sliding clamp of the replisome, gp45, which is loaded onto the DNA at single-strand interruptions by a gp44–gp62 clamp loader complex (158–160, 349, 350) (see also “DNA Replication” below). This sliding clamp tracks along DNA and allows RNA polymerase, bound at late promoters, to form open complexes, thereby coupling late T4 transcription to DNA replication. The α-subunits of the core polymerase play distinct additional roles in recognition of late promoters (402). In certain recombination-and-ligase-defective mutants, late transcription can occur without DNA replication, because single-strand interruptions in the DNA provide entry sites for the sliding clamp (129).

Genes downstream of any promoter may be poorly expressed because of transcription termination, because of RNA processing or degradation, or because ribosome binding sites are sequestered by secondary structures. Certain *E. coli rho* (transcription termination factor) mutations (also called host defective, *hdF* or *nusD*) prevent growth of wild-type T4 by causing premature termination of many T4 transcripts (383, 384). Mutations that allow growth in these *rho* mutants have been found in three nonessential T4 genes. It was initially thought that these encode transcriptional antiterminators, but current evidence suggests that at least one of them, *goF*1, allows T4 growth in the *(nusD) rho* mutants, because it stabilizes the few functional transcripts that have not been prematurely terminated. Plasmid-encoded Rop protein has a similar stabilizing effect, and pBR322 plasmids allow T4 growth in these *rho* mutant hosts (379, 422).

The transitions from early to middle to late transcription are influenced by many factors, each individually having only minor effects (209). DNA topoisomerase mutations affecting torsional stress in the DNA specifically affect late transcription patterns (240). Moreover, a DNA binding protein (DsbA) may affect transcription from some, but not all, late promoters (146). Several nonessential T4 proteins affect the binding of minor host σ factors to core RNA polymerase, preventing competition with σ^{70}, the only host sigma factor that is essential for T4 development. For example, the Mrh protein modulates phosphorylation of the host’s σ^{32}, which affects T4 late transcrption at 42 °C (121, 122, 288), and the Srh and Srd proteins are proposed to be decoys of the host’s σ^{32} and σ^{70}, respectively (288).

The late genes code for virion components, for some DNA repair and recombination proteins, and for proteins that cut and package the complex vegetative DNA into preformed heads. A soluble lysozyme (gp e) lyses the cell wall of host bacteria to release the progeny phage particles. It is evolutionarily related to a baseplate lysozyme (gp5) that attacks the cell wall from the outside during infection (81, 189, 292). The gp e lysozyme gains access to its substrate in the outer cell wall through a protein channel that forms from the holin (gp t), when the membrane potential breaks down. In T4 this process is regulated, in part, in response to superinfection, sensed by the gp rI (1, 315, 329, 330).

Post-Transcriptional Controls

As with other phages, post-transcriptional mechanisms modulate expression of T4 genes in a variety of ways: (i) by modulating translation initiation via potential secondary RNA structures and/or by coupling translation of two or more genes, (ii) by translational repression, (iii) by proteins affecting ribosome structure and function, (iv) by processing of primary transcripts, and (v) by the rate of degradation. In addition, expression of at least four T4 genes requires splicing, or in a variation so far unique among phages, skipping of segments of primary transcripts.

RNA Structures

Due to the interspersion of early and late genes on the T4 map (figure 18-2), many early and middle transcripts are extended into late genes (figure 18-4). Nevertheless, few corresponding late proteins are synthesized early at temperatures below 37 °C. It is remarkable that expression of at least 10 such late genes investigated so far follows similar patterns: in the long early transcripts a hairpin sequesters the translation initiation region, either the Shine–Dalgarno sequence or the initiation codon, or both. A late promoter immediately upstream of the late gene(s) directs synthesis of transcripts that cannot form the hairpin; these transcripts are efficiently translated (figure 18-4). Expression of gene *49* coding for endonuclease VII (18) is one example. These late genes can be translated early in vivo at high temperatures, or in vitro when the RNA is broken, because in broken early RNA some sequestered late ribosome binding sites become accessible. We speculate that this pattern may have selective advantage because if RNA polymerase pauses at certain hairpins formed within early transcripts, then transcription initiation from an adjacent late promoter might be facilitated.

Translational coupling can coordinate synthesis of proteins encoded in multicistronic RNAs. It has been implicated in T4 gene expression early on (380). It can serve to coordinate expression of proteins involved in the same pathway (e.g., the tail fibers gp34 and gp35, or subunits of the terminase). Translation of the terminase subunit gp16 may disrupt a hairpin that otherwise prevents translation initiation of the overlapping gp17 (119). Translational coupling can also adjust the synthesis of proteins that form heteromeric complexes. For example, the pentameric clamp loader complex contains four gp44s and one gp62 (133, 403, 440). This ratio is regulated by translational coupling. In addition, coupling might interconnect different physiological processes, for example coupled translation of *vs* (modifier of valine synthetase) and *tk* (thymidine kinase) might allow communication between translation and DNA metabolism.

Translational Repressors

Three regulatory systems of T4 depend on translational repression that is autogenously regulated. Gene *32* codes for the major single-stranded DNA binding protein (SSB) involved in DNA replication, recombination and repair (10). Its expression is controlled by a pseudoknot upstream of the ribosome binding site and adjacent A-rich repeats (257, 261, 360), by cooperative binding to single-stranded (ss) DNA and RNA (412, 415), which is stronger to ssDNA than to RNA. Binding of gp32 to its mRNA is nucleated by preferential binding to the pseudoknot. Cooperative binding of gp32 to the adjacent unstructured leader transcript eventually obliterates the ribosome binding site and prevents more synthesis of gp32 when, and only when, there is excessive gp32 in infected cells (99, 220, 222, 257, 327, 360–362, 415). This beautiful, intricate system allows protection of all ssDNA without preventing general translation from most mRNAs under many different growth conditions.

Gene *43*, encoding DNA polymerase, is similarly autoregulated (222). A stem-loop in the leader RNA binds excessive DNA polymerase, thereby obliterating the adjacent ribosome binding site (12, 261, 323, 324, 418).

A more general, nonessential translational repressor, RegA protein, binds to several specific transcripts, reducing translation of several T4 replication proteins and of some host proteins. This repression is most apparent under experimental conditions that prolong early transcription (429). The RegA protein binds to AU-rich regions of several leader regions, but a consensus sequence is not apparent. The crystal structure of this protein is known (188), and interactions with target RNA are inferred from mutational studies (3, 4, 139, 238, 261, 263, 312, 326, 354, 404, 405, 424, 434).

T4-Evoked Modifications of Ribosomes

Translation of host messages ceases immediately after T4 infection. Mysteriously, it is inhibited by infection with DNA-free particles, so-called ghosts (100). It is thought that some cells can recover from this immediate turn off, but in successfully infected cells, ribosomes have already been programmed to translate T4 messages exclusively. A few host membrane proteins are the notable exception. There is extensive, controversial literature, reviewed with great fairness (428), concerning to what extent T4 infection alters the host's ribosomes. On balance, it appears that ribosomes from T4-infected cells differ mainly in the content of their S1 protein. This protein is one of several T4 proteins that are ADP-ribosylated by the T4 modB protein (Wolfgang Rüger, personal communication). It is conceivable that this modification lowers the affinity of S1 for ribosomes, and that ADP-ribosylation by the related Alt protein, which is packaged into phage particles, is responsible for the ghost effect mentioned above. It is suspected but not yet shown that ribosomes acquire a few phage-encoded proteins after infection (227).

RNA Processing and Degradation

Host RNases, especially the RNase E complex (69, 144, 298, 410) now called the "degradosome" (see (391) for review) and Rnase III (75, 88), contribute to processing of T4-encoded tRNAs (352) and to degradation of T4 mRNAs (69, 281, 291). In at least one case RNase III cleavage activates translation from an internal ribosome binding site to produce a truncated terminase protein, gp17‴ (119).

Selective degradation of early T4 transcripts at late times is directed by at least two mechanisms. A T4 *regB*-encoded ribonuclease cleaves early transcripts in their ribosome binding sites (72, 346, 347, 351, 406). Other yet unknown nucleases, indirectly controlled by MotA, are thought to selectively degrade early transcripts. They spare late transcripts because the T4 *dmd* gene, among others, directly or indirectly protects late transcripts (184–186).

RNA processing is of utmost importance for the final activities of eight T4-encoded tRNAs (reviewed by (352) and (281)). Although these tRNAs are nonessential under standard laboratory conditions, they are essential under other conditions and are thought to enhance efficient synthesis of most T4 gene products. The codons corresponding to the T4 tRNAs are used significantly more frequently in T4 than in *E. coli*. The T4 tRNA region can be transcribed from several promoters (56) as large transcripts containing all genes and from several internal promoters. The stepwise processing yielding eight tRNAs and two small RNAs of unknown function is accomplished by host RNases, to some extent by autocatalysis, and by a nucleotidyl transferase that adds CAA to some of the products (88). At least one nonessential T4 gene, *mb* = *M1* = *cef*, is important for processing of three tRNAs. This gene becomes essential when a host gene, probably coding for a tRNA, is defective (340).

Introns

The genes *td, nrdB, nrdD*, and *60* are interrupted by intervening untranslated sequences (24, 171, 369). The first three of these introns can self excise (24, 369). They form three-dimensional structures related to those of other type I introns (24, 369). The *td* and *nrdD* introns encode endonucleases of the GIY-YIG type, wheras the *nrdB* intron encodes a defective endonuclease of the H-N-H type. In the related RB5 phage this intron endonuclease is complete. These endonucleases are important for "homing" of the introns (24). Free-standing, related endonucleases *segA* through *segF* (363) and *mobA* through *mobE* (262) exist in the T4 genome, but are lacking in many related genomes. The Seg endonucleases (and probably the Mob enzymes as well) introduce sequence specific DNA breaks (28, 182, 183, 364, 365) whose recombinational repair leads to exclusion and gene conversion to alleles of the unbroken chromosome of markers close to the break (24, 28, 290, 300, 365) (G. Mosig, unpublished data). Actually, the SegF endonucleae appears to be a chimera of Seg and Mob-like endonucleases (128, 262).

In contrast, the intron in gene *60* is not excised but is skipped during translation, probably because it can fold into a superstable hairpin that is skipped by the ribosome (134, 171).

Protein Degradation

In host strains containing the *e14* element, a defective prophage, the host-encoded translation factor EF Tu is cleaved and functionally inactivated for translation of late T4 proteins (132, 377, 442). An N-terminal peptide of the major capsid protein gp23, called Gol, combines with the Lit protein, encoded by the host's *e14* element, to cleave EF Tu at a specific site. Because gp23 is a late protein, only late translation is affected.

Modification and Restriction of T4 DNA

In the following discussion the term "restriction" is used in its broadest meaning and is not limited to type II restriction enzymes. The complex modifications and restrictions of T4 DNA and of other DNA by T4 can be best rationalized as a result of ongoing evolution that is accelerated by strong selective forces and exchanges between the genomes of T4 and its host, including plasmid and prophage genomes that reside in the host.

As mentioned above, T4 DNA contains HMC instead of C residues. This modification protects the T4 DNA against most foreign restriction enzymes and against T4-encoded restriction endonuclease II (gp denA). The latter enzyme, together with endonuclease IV (gp denB) and the gene *46/47*-controlled nuclease, degrade the host DNA as part of the parasitic strategy to usurp the host (66–68). Although HMC residues confer resistance to most restriction systems, they render DNA susceptible to the Mcr restriction systems of the host. These host functions were the first restriction enzymes discovered (247). Originally called rgl they are now called McrA and McrBC, because they restrict DNA containing methylcytosine or HMC.

McrA of *E. coli* K-12, like the *lit* gene mentioned above, resides in the *e14* element (17). Thus, in *e14*-containing bacteria T4 is excluded at several levels.

The *mcrBC* genes, probably acquired by a recent transposition, are part of a cluster of restriction genes at 99 min on the *E. coli* K-12 map (68). The proteins assemble into hexameric GTP-dependent rings (317). A shorter version of McrB, McrB$_s$, modulates McrB's activity (318).

These restriction functions are inactive when the HMC residues are glycosylated. In T4 DNA, all HMC residues are modified: 70% with α-, and 30% with β-glucosyl linkages. In addition, an early T4 antirestriction protein (gp arn) protects nonglucosylated T4 DNA against McrBC but not other host restriction enzymes.

T2, T4, and T6 encode α-glucosyl transferases, which modify HMC DNA after it has been polymerized. In T2 and T6 DNA, there are no β-glucosyl transferases, and 25% of the HMC residues remain unglucosylated, probably because the α-glucosyl transferases cannot modify two adjacent HMC residues. T6 contains many di-glucosylated residues. The relevant enzyme is unknown.

In addition, T2 and T4, but not T6, encode a Dam methylase that methylates 0.5–1.5% of the adenine residues at the N^6 positions, mostly but not exclusively at GATC sequences. These enzymes exhibit patches of similar amino acids with the *E. coli* Dam methylase and the *Dpn*II methylase of *Diplococcus pneumoniae*. The only proven physiological role of adenine methylation is protection against the phage P1 restriction system (see chapter 24), when there is no HMC glycosylation (68).

The classical example of phage exclusions by genes of resident prophages is that of T4 *rII* mutants by the *rexA* and *rexB* genes of λ lysogens. This exclusion was elegantly used in Benzer's classical analyses of structure and function of a gene. It occurs at the time of transition from join-copy to join-cut-copy recombination, and it involves several enzymes important in the latter mechanism (279, 296). The molecular mechanism of this exclusion is unknown. It has been proposed that the RexB proteins form an ion channel that, when opened after infection with T4 *rII* mutants, leads to cessation of host metabolism (321).

Another cryptic DNA element of certain *E. coli* strains, *prr*, encodes a protein that excludes T4 RNA ligase 1/polynucleotide kinase deficient mutants. PrrC protein is a cryptic RNA endonuclease that is activated by the small (26 residues) T4 Stp protein to cleave the anticodon loop of an essential host $tRNA^{Lys}$. T4 RNA ligase 1/polynucleotide kinase can repair this damage, but in the absence of this T4 protein the cleavage of this tRNA is lethal to T4 protein synthesis. Intriguingly, the *prrC* gene is located between three genes of a type IC restriction cassette.

Phage P2 lysogens exclude T4 by two mechanisms: the P2 Tin protein poisons the single-stranded DNA binding protein gp32 that is essential for all T4 DNA replication and recombination (297), and the P2 Old protein can degrade T4 (as well as λ) DNA from ends, nicks, and gaps (63), unless the ends are protected by certain proteins (gp2 in the case of T4; 14, 244, 419).

DNA Replication, Recombination, and Repair

T4 codes for all the components of its own replication and recombination complexes, for an excision-repair enzyme, denV, and for many enzymes that synthesize precursors for or modify T4 DNA. Some of the latter proteins (dCTPase, dCMP hydroxymethylase, deoxynucleotide kinase) specifically allow incorporation of HMC triphosphates during T4 DNA synthesis. The α and β DNA glucosylases sugar-coat the HMC-containing DNA after it has been polymerized, to protect it from attack by Mrc nucleases described above.

The basic replisome proteins of all organisms share sequence and structural similarities, and some of the T4 replication proteins can partially function in eukaryotic in vitro systems (307).

Initiation of T4 DNA replication is far more complex than structure and function of the basic replisomes. As we discuss in detail below, replication forks can be initiated in several different ways. There is necessary, but limited initiation from bonafide origins; most subsequent replication forks are initiated from intermediates of homologous recombination. There are several redundant origins of DNA replication, and there are several pathways of recombination-dependent DNA replication. We surmise that these redundant pathways are exquisitely suited to adjust DNA replication to different transcription patterns, described above, to environmental conditions and to packaging of DNA described below.

The Basic Replisome

The combination of genetic experimentation and virtuoso biochemical and biophysical characterization of replication proteins has led to an understanding of functions and interactions of replication proteins in the basic replisome, a biological machine that moves the replication fork (or through which replicating DNA is passaged) (307, 310) (figure 18-5).

Seven T4 proteins, corresponding to genes *43* (DNA polymerase), *44* and *62* (sliding clamp loader), *45* (sliding clamp), *41* (DNA helicase), *61* (primase to synthesize primers for Okazaki fragments), and *32* (SSB) together constitute the basic replisome, a biological machine (7, 8) that can move the replication fork through model templates at in vivo speeds. To seal Okazaki fragments, RNase H has to remove the RNA primers (and adjacent DNA) and DNA polymerase has to fill the gaps so that DNA ligase can ligate the DNA fragments (29, 40, 74, 179, 180, 301, 309). Additionally, T4 gp59 is essential for loading this helicase to recombination junctions in vivo (286). It also helps load the gp41 helicase to an origin in vitro (179, 309). In vivo host DNA polymerase I or host RNaseH can substitute for T4 RNase H (169), and *E. coli* ligase can substitute for T4 ligase, if T4 *rII* genes are mutated (34, 73, 191, 221). The three-dimensional structure of several of these proteins (or segments of them) (301, 302, 357–359) can now be correlated with defects defined by mutations. Assembly and interactions between these proteins are being elucidated by sensitive biophysical methods (29).

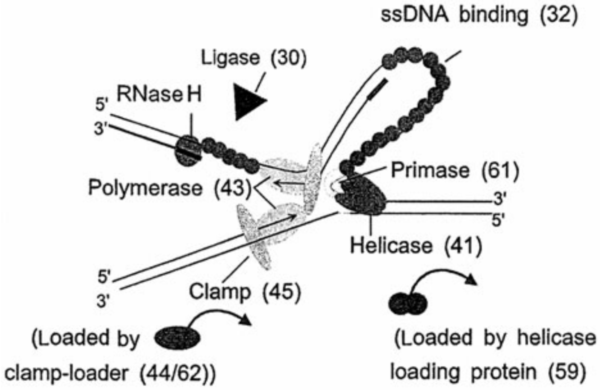


Figure 18-5 The proteins of the DNA replication fork. From N. Nossal, with permission.

In contrast to the *E. coli* replisome, the T4 replisome has not been isolated as a functional complex. Possibly, the T4 core replisome components interact less strongly than those of other organisms to facilitate interactions of DNA polymerase with different accessory replication and recombination proteins during different developmental stages as discussed below. Clearly, a real understanding of the intracellular activities of individual enzymes in such complexes requires a combination of mutational, biochemical, and biophysical analyses.

This has been done exquisitely for the DNA polymerase of T4, and the related RB69 DNA polymerase (20, 120, 193, 359, 420, 421). Mutator and antimutator mutations occur in domains that are important for polymerizing and proofreading activities. These studies also revealed hinge regions that are important for switching the primer terminus between the polymerizing and proofreading site. All these contribute to the exceptionally high fidelity of this enzyme (95, 97, 308, 334–337).

Similarly, mutations affecting the core of the SSB, gp32 (357, 358), define some interactions with other proteins (279, 290). In addition, limited proteolysis of this protein has revealed the importance of the N-terminal segment for cooperative binding to DNA and of the C-terminal segment for interactions with several recombination proteins (433).

Some DNA replication genes are clustered, but others are scattered throughout the genome. Most are transcribed from both early and middle promoters. Coordination of T4 DNA replication with other processes is achieved by the assembly and disassembly of protein complexes that use stretches of DNA covered with SSB, gp32, as scaffolds. These interactions have been characterized in vivo and in vitro (115, 270, 274, 283, 284, 296, 427).

In wild-type T4, leading- and lagging-strand syntheses are coupled, but they can be uncoupled both in vivo and in vitro (181, 293). Primase mutants are viable because at T4 origins RNA polymerase-dependent transcripts prime leading-strand DNA synthesis (246, 287), and priming of Okazaki fragments can be bypassed by recombinational mechanisms (290, 293) discussed below. In topoisomerase mutants, leading- and lagging-strand synthesis are also partially uncoupled (240, 401), and their replication also depends on recombination (111, 173, 212, 279, 290, 306, 387).

Initiation from Origins

Several origins of replication have been detected under various conditions (147, 202, 215, 216, 240, 248, 252, 282, 385, 439); for reviews see (218, 287). In infecting T4 chromosomes, the first round of DNA replication is initiated from only one of these origins (85, 86), perhaps because at the time of origin initiation there are limited supplies of replisome components. Because of the circular permutation of chromosomal ends (indicated in figure 18-4), in each individual chromosome any origin is located at a different distance from the DNA termini. Four of these origins (*A, E, F,* and *G* in figure 18-2) have been closely investigated. Each one has a different sequence. Three origins (*A, F,* and *G*) require transcription from middle promoters (143, 218, 248, 286, 287), whereas *oriE* uses an early promoter (287, 407). For simplicity, figure 18-4 depicts initiation from an arbitrary generic origin. Transcripts initiated from the origin promoters serve as primers for leading-strand DNA synthesis. These transcripts also code for proteins, but obviously priming (requiring transcripts to reinvade the DNA) and use as mRNAs (requiring access of ribosomes) are mutually exclusive.

In vivo, the transition from RNA primers to leading-strand DNA occurs at several sites more than 1 kb downstream of the primer promoter (286, 287, 407), implying that the priming transcript has to reinvade the DNA, forming an R-loop. Reinvasion can be facilitated by global torsional stress in the DNA. Indeed, in vitro initiation from *oriF* has been achieved with an RNA primer that had been hybridized to supercoiled plasmid DNA (309). The 3′ end of this RNA, located at the end of an unwound region, determined the priming site in vitro. This site is different from the predominant in vivo priming sites, determined by primer extension, near the transcription terminator between *uvsY.-1* and *uvsW* (287).

In contrast, *oriE* (147, 211) can function when torsional stress is reduced (e.g., by excessive ^{32}P damage, by oxidative stress damage, or by mutation in the host gyrase gene; 240, 294). *OriE* function depends on a small protein, RepEB, and the auxiliary protein RepEA, both encoded by sisters of the primer transcript, and on at least one of five repeat sequences, so-called iterons upstream of the primer promoter (287, 407). The binding of RepEB to one or more of these iterons is required for *oriE* to function (150). It is postulated that RepEB binding to an iteron facilitates loading of a DNA helicase upstream of the primer promoter, and that unwinding of DNA behind the RNA polymerase by the helicase can compensate for the lack of global torsional stress in *oriE*. In some respects *oriE* resembles *oriC* of *E. coli*, in which leading-strand DNA synthesis can alternatively be initiated from primase-dependent RNA or from primer transcripts (210). Less well characterized *oriK*s also depend on priming by transcripts (208).

Several other T4 origins (202, 439) have not been mapped precisely. Interestingly, an origin near gene *17* has been detected only in nonmodified (cytosine-containing) DNA (439), or when the host contained a mutation in the transcription terminator *rho* gene (385). Possibly altered transcription termination on cytosine-containing DNA, by T4 gp alc or by certain *rho* mutations, may be important for activating this origin.

The transition from σ^{70}-dependent prereplicative transcription to T4 σ^{gp55}-dependent late transcription inhibits initiation from these replication origins because the modified RNA polymerase no longer recognizes origin promoters (246, 407) and because a late protein, the UvsW RNA–DNA helicase, actively unwinds any persisting R-loops and prevents transcripts from serving as primers for DNA synthesis (65, 102).

Recombination and Recombination-Dependent DNA Replication

The detection of homologous recombination has provided a powerful argument that phages are models for living systems (165). A strong connection between replication and recombination was suspected early on (413). This hypothesis fell temporarily out of favor when it was shown that in phage λ viable recombinants and in T4 branched recombinational intermediates can be formed in the absence of DNA replication (57–59, 258). However, the recombination and segregation patterns at the ends of incomplete genomes (packaged in small T4 particles, see below) (276, 289) implied and led to the first compelling demonstration of recombination-dependent DNA replication (246). The current view of these processes has been thoroughly reviewed (277, 279, 290).

As soon as the first replication forks have reached an end, recombination intermediates formed at these ends compete effectively for assembly of recombination-dependent replisomes by what we have called join-copy recombination/replication (pathway II in figure 18-4). Because the ends correspond to random positions of the genetic map, origins are no longer preferentially replicated (85, 246).

The terminal redundancies of individual T4 chromosomes are sufficiently large (~ 5 kb) to allow recombination between replicated and unreplicated termini after infection of *E. coli* by a single T4 particle (280). In electron micrographs, the resulting structures appear like rolling circles or more complex "firewheels." Firewheels as intermediates of T4 DNA replication had been proposed earlier, but they had been interpreted in terms of rolling circle initiation (425, 426). However, later experiments showed that they are formed by multiple recombinational invasions (85).

After late proteins have been synthesized, a late "join-cut-copy" recombination/replication pathway can operate (figure 18-4). These and additional pathways are discussed below. Keeping in mind that there are many ambiguities and caveats in defining recombination pathways based on participating enzymes or structures of recombining DNA, five different recombination/replication pathways can be distinguished (279). They are not rigorously separated, but contain some common steps, and depend on some common proteins that are mixed and matched with other proteins in different complexes directing different pathways.

Obviously, the time when any pathway is active depends on expression of the genes required for that pathway. Because late gene expression depends on DNA replication, and because some proteins participate both in packaging of DNA and in recombination, all these DNA transactions are tightly interwoven (figure 18-4). Redundant alternative modes of replication and recombination ensure that these processes work under many different conditions and during different stages of development (290, 293).

The single-strand annealing pathway I (57, 58) was analyzed when DNA replication was prevented. Under this condition the infecting DNA has to be broken, or nicked, and ssDNA regions have to be generated by nucleases before complementary DNA strands from different parents can anneal. Therefore, recombinational intermediates appeared only late after infection, and no viable recombinants were produced. Under replication-proficient conditions recombinogenic ssDNA termini appear earlier, due to failure to initiate an Okazaki fragment at each terminus. If they anneal with a homologous region, replication can also be initiated from their 3′ end (279) (figure 18-4). Gp32 can catalyze annealing, explaining the viability of all T4 *uvsX* mutants (341). The DNA branches generated by ssDNA annealing are obviously less complex than D-loop branches, and therefore their resolution may depend less on endonuclease VII (89), which is actually inhibited by UvsX and UvsY proteins (47). Both annealing and invasion are also important to initiate recombination in phage λ and yeast (319, 381), and probably in all organisms.

In the presence of UvsX and UvsY proteins, the unreplicated ssDNA termini invade double-stranded homologous DNA to form D-loops. T4 UvsX is a RecA homolog (264) and T4 UvsY is a hexameric recombination-replication mediator that facilitates loading of UvsX to ssDNA covered with gp32 (22, 23, 51). Gp59, another mediator protein, facilitates loading of the gp41 helicase to recombination junctions, a process that is important to establishing replication forks and also to driving branch migration of recombining DNA (179, 180, 301, 348). Replication initiated from the invading DNA of D-loops, join-copy replication/recombination (246), is depicted as pathway II in figure 18-4. It occurs early because it requires only prereplicative T4 proteins.

In contrast, late recombination pathway III (figure 18-4) requires, in addition, endonuclease VII or terminase (290, 293). When these enzymes make a single cut in the invaded strand of the D-loop, the 3′ end can initiate replication of that parent whose ends invaded the D-loop. Because endoVII and terminase are synthesized predominantly late, pathway III is a late pathway. We have dubbed this pathway "join-cut-copy" (277, 290). It is essential in primase mutants (and important in topoisomerase mutants) so that the DNA strand that was displaced during origin-initiated replication can be copied in primase and topoisomerase mutants (figure 18-4). Origin-initiation of these mutants starts at the same time as in wild-type T4, but recombination-dependent replication is delayed compared with wild-type T4, causing the so-called DNA-delay (DD) phenotype (279, 290, 293).

The classical double-strand DNA break repair pathway IV (393) results in only limited DNA synthesis. This pathway and a synthesis-dependent strand annealing pathway V (305), based on in vitro results with T4 proteins (114), have been proposed to explain homing of T4 introns (76, 299, 300) and the repair of double-strand breaks introduced by endonucleases that resemble intron nucleases (28). This repair also uses multiple pathways (24, 300, 320).

The in vitro recombination-dependent replication reactions (114, 213, 214) are primed by 3′ ends of ssDNA. With an excess of UvsX protein, and in the absence of gp uvsX, primase or topoisomerase, they result in a conservative DNA replication called "bubble-migration" in which the nascent DNA strand is extruded from the DNA template much like nascent transcripts are extruded by RNA polymerase (114). There is no direct evidence that bubble migration occurs in vivo at physiological concentrations of UvsX protein and in the presence of topoisomerase (218). Surprisingly, the T4 proteins gp46 and gp47, whose genes are homologous to the eukaryotic Mre11-Rad50 complex (79, 80) and yield the most recombination defective mutations (35, 37, 38), are not required in the current in vitro recombination systems.

Ultimately, reiteration of recombination-dependent DNA replication by these different pathways generates a highly branched concatemeric DNA network in which no individual chromosomes can be distinguished (172). The branches can be resolved either by endo VII (gp49) or the largest endonuclease subunit of terminase (gp17). Endonuclease VII (gp49) cuts Holliday and 3-way recombinational junctions as well as heteroduplex mismatches and other DNA distortions in vivo and in vitro (43, 44, 46, 112, 123, 137, 141, 198–200, 204, 295, 296, 328, 367, 378). Gp17 of terminase can cut single-stranded regions in supercoiled DNA (41, 42) and at junctions between ssDNA and dsDNA (118). The endonuclease activities of T4 terminase, the enzyme that generates chromosomal ends during packaging, reside in the subunits encoded by gene *17* (41, 42, 223, 234, 242, 250). At least four proteins are initiated from different in-frame initiation codons. Only the largest of these has a ssDNA binding domain that allows it to bind to and cut preferentially the junctions of ssDNA and dsDNA (118). It is surmised that only the largest gp17 is also involved in join-cut-copy recombination-dependent DNA replication (290). In any case, because only the join-cut-copy mode can bypass the requirement for primase or topoisomerase in T4 DNA replication, double and triple mutants defective in these enzymes are defective in the bypass replication (290, 293).

Extensive heteroduplex regions containing mismatched or looped-out bases are formed during recombination in vivo (295, 296) and in vitro (45, 46, 198). Their repair is mediated by endo VII, the enzyme that cuts Holliday junctions (198, 200, 296), not by mutSHL or its homologs (97). Interestingly, this repair appears to occur during packaging, when endoVII is associated with the portal protein, gp20 (138).

Recombination frequencies decline rapidly when the recombining molecules share less than 50 bp of homology (19, 94, 136, 371). This decline is probably not due to inability to pair, because in vivo and in vitro heteroduplexes containing extensive mismatches are formed (45, 295, 296) and the UvsX-related *E. coli* RecA protein requires only a few homologous base pairs to initiate pairing (170). Instead, recombination seems to be reduced by an enhanced probability of cutting by endo VII during branch migration through partially heterologous regions and repair of mismatched heteroduplex regions (see below) (46, 123, 141, 198, 200, 204, 290, 295, 296, 366, 367, 378, 441).

DNA Repair

As in other organisms, damage and mismatches in T4 DNA can be repaired in several different ways, and repair defects result in increased sensitivities to such damage and/or to differences in mutation rates. Recombination-dependent DNA replication is an important component of repairing DNA damage and of restoring broken or stalled replication forks (37, 38, 82, 131, 205, 213, 214, 217, 219, 271, 280, 387). There is, however, a difference between recombination patterns at "natural" chromosomal ends and secondary ends generated by endonucleases. All markers near natural chromosomal ends appear in progeny particles (276, 289). In contrast, some genetic markers near nuclease-generated ends are degraded in both DNA strands from the broken chromosome and are replaced by alleles from the unbroken chromosome, resulting in apparent gene conversion (24, 128, 365, 386). We surmise that T4 gp2 (370) protects natural, but not secondary, chromosomal ends from degradation.

In addition to the roles of homologous recombination pathways in DNA repair, discussed in the preceding section, there are two nonrecombinational repair pathways. The first mechanism shown to repair UV-damaged T-even DNA was photoreactivation (103). It is now known that the host enzyme responsible for this repair reverts exclusively pyrimidine dimers to the monomeric state (reviewed by 96, 97).

The second mechanism was first revealed by Harm (149), who isolated DNA repair mutants in two T4 genes that are now called *denV* and *uvsX*. These mutations defined excision repair and recombinational repair as major T4 pathways. These pathways are physiologically connected, because the ssDNA binding protein, gp32, participates in both excision repair and recombinational repair (271) and because the intermediates of excision repair are thought to generate substrates for recombination, for example by stalling or disrupting replication forks (245).

The product of *denV*, endonuclease V, is the prototype of a base excision repair protein. It has both *N*-glycosylase and abasic lyase activities (90, 245), which incise the DNA (forming a covalently linked protein–DNA intermediate). These activities remove the pyrimidine dimers to allow resynthesis, notably including *E. coli* DNA polymerase 1 (392). Profound differences in UV sensitivities of T4 and T2 are due to the presence of *denV* in T4 but not in T2 (90, 149).

The precise role of the gp46/gp47 complex in DNA repair is still as elusive as it is in recombination. A suppressor of gene *46/47* defects, *das* (157), suppresses the DNA repair deficiency but not the DNA synthesis arrest phenotype of T4 uvsX mutations (417). *Das* mutations change the sequence of T4 RNaseH without inactivating this enzyme (G. Mosig and N. Colowick, unpublished data), suggesting that they allow RNase H to substitute for defective nuclease activities of the 46/47 complex.

DNA Packaging

The branched, concatemeric vegetative T4 DNA generated by recombination-dependent DNA replication is cut by terminase, a heteromeric protein encoded by genes *16* and *17* (331, 333). This protein complex binds to the vegetative

DNA and subsequently connects with gp20, the portal vertex protein of the head, to form a "packasome" that, fueled by ATP hydrolysis, drives DNA into preformed heads (50, 241, 243). Gp16 binds to dsDNA. The endonuclease activity and a weak ATPase activity of T4 terminase reside in gp17 (41, 42, 119, 234, 332). Gene *17* produces several proteins of different sizes by initiation from in-frame internal initiation codons (119). These internal initiation sites are preferentially used in transcripts initiated from an internal promoter and in transcripts processed by RNase III (119). As predicted from the similarity of its N-terminal domain with ssDNA binding proteins, the largest of these, and only the largest, binds to ssDNA segments and cuts them preferentially at junctions with dsDNA (118). Presumably, this facilitates initiation of packaging from recombinational junctions. The shorter products of gene 17, lacking the ssDNA binding domain, suffice for packaging of precut mature DNA in vitro (138).

Recombinational branches can be cut by T4 endonuclease VII (gp49), discussed above. Alternatively they are resolved by branch migration or by DNA synthesis (296), explaining why gene *49* mutations are not completely defective in recombination. DNA ligase, endonuclease V, and topoisomerase are also required, presumably because only uninterrupted dsDNA can be wheeled into heads by the packasome (48). Because, as mentioned above, endonuclease VII and gp17 also play important roles in late, join-cut-copy recombination-dependent DNA replication, these activities link DNA replication and packaging at a physiological level.

The initiation of DNA packaging is a matter of debate, suggesting that this process, like other DNA transactions, can be initiated by different redundant mechanisms. A mechanism based on synapsis of two homologous sequences in genes *16* and *19* and on recognition of this sequence as a *pac* site by gp16 (243) has been proposed (48). This model is based on preferential amplification of a DNA segment encompassing gene *17* between the two short homologous sequences (437, 438), on preferential packaging of foreign DNA containing this sequence, and on the apparent overabundance of the corresponding restriction fragment in packaged T4 DNA.

On the other hand, comparisons of maps based on cutting of chromosomal ends versus recombination showed no distortion in the gene *16* to gene *19* region, and only a slight distortion near gene *37* (273, 278, 289) (figure 18-2). Because the largest peptide made by gene *17* binds to ssDNA, and not to dsDNA, and cuts DNA preferentially at ssDNA–dsDNA junctions (118), we have proposed that packaging is preferentially initiated at recombination junctions, which occur at random positions of the genome.

Probably both mechanisms can initiate packaging, and the different results reflect differences in the DNA substrates. In contrast to the experiments of Mosig (273, 278, 289), who investigated wild-type phages containing modified HMC DNA, the substrate for the experiments of Lin and Black (241) was unmodified dC-containing DNA since modified T4 DNA would not be cut by restriction enzymes. As mentioned above, the gene *17* region also contains a replication origin that is preferentially used only in dC-containing DNA (439). Processive packaging of 103–105% genome lengths to fill the preformed heads would randomize the ends in either case.

We consider the possibility that the basic terminase of a T-even ancestor was composed of gp16 and gp17′, recognizing a preferred *pac* site. Later acquisition of a ssDNA binding domain allowed the T4 terminase to better adapt packaging to the recombination-dominated life-style, which is best served by random permutation of its chromosomal ends.

Virion Structure and Assembly

It is now clear that in all phage assembly systems studied, obligatory pathways of morphogenesis are determined at the level of protein–protein interactions and not at the level of sequential transcription of the structural genes (107). However, the complex regulation of T4 transcription and translation, discussed above, modulates the final levels of structural gene products late in infection. These regulated levels, in turn, are important in controlling successful assembly of the phage structures (108). The major components of T4 virions, (heads, tails, and long tail fibers) are assembled via independent pathways in vivo. These substructures can be isolated and combined in vitro to yield infectious particles (105, 108).

Advances in cryo-electron microscopy and X-ray diffraction now provide structural data on bacteriophages from below 2 nm to atomic resolution (for reviews see 16, 344, 345). Recent cryo-electron microscopic studies of T4 isometric heads (176, 311) confirm and extend earlier head-structure data, and combined crystallographic, and data on T4 tails (113, 187) provide structural details that are relevant to the understanding of T4 tail function.

Head Structure and Assembly

The head assembly pathway is summarized in figure 18-6. The completed T4 head is composed primarily of the major cleaved capsid protein, gp23*, arranged in an icosahedral surface lattice of $T = 13$, elongated along a 5-fold axis (267). The length of the T4 head is defined by a number, Q, related to the icosahedral T number. An illustration of Q numbers is shown in figure 18-7 together with models of some T4 head lengths. The 5-fold vertices, except for one, are occupied by the cleaved vertex protein (gp24*), whose sequence is related to gp23, probably reflecting gene duplication. The 5-fold vertices are likely replaced by gp23* in certain gene 23 missense mutants that bypass the requirement

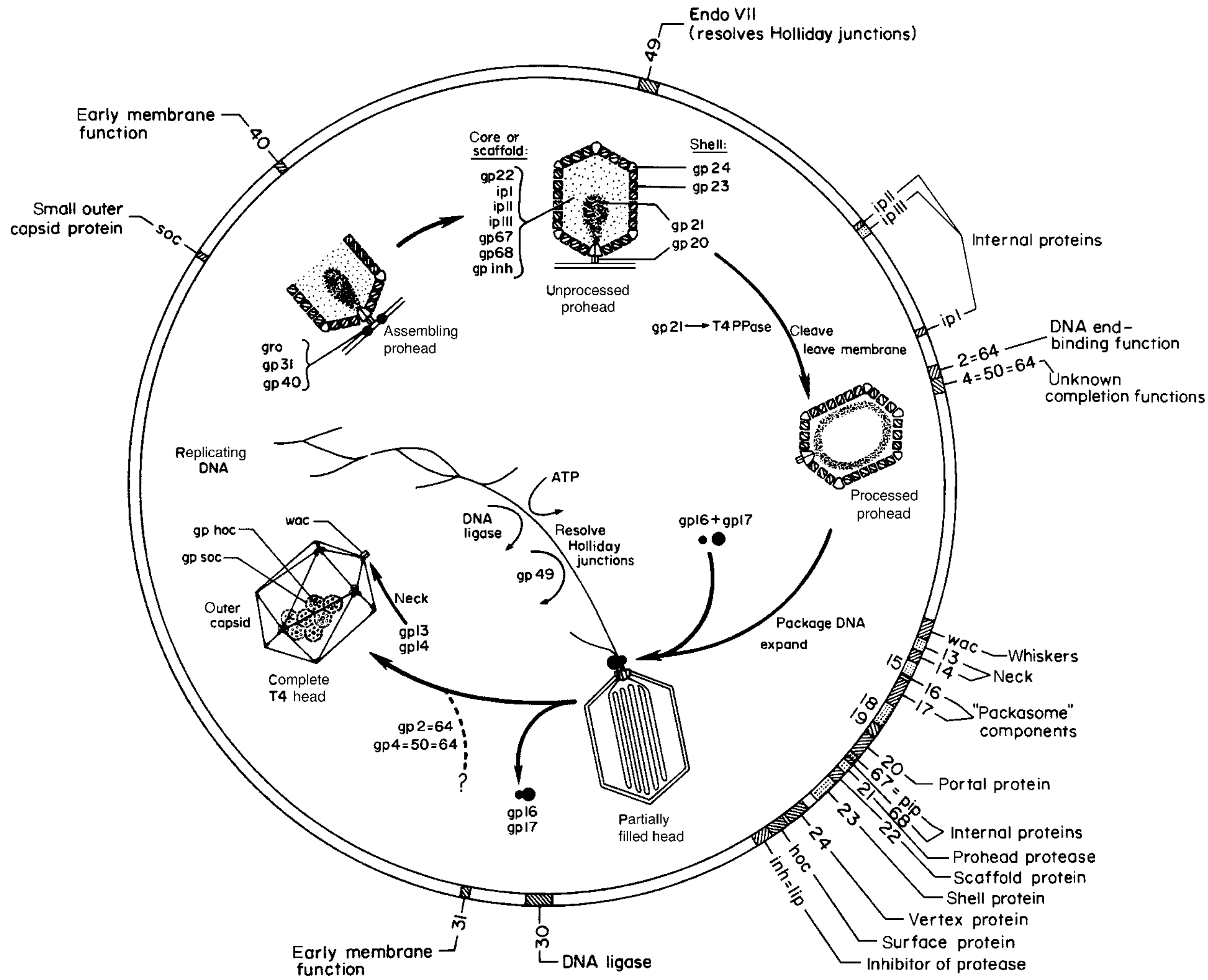

Endo VII
(resolves Holliday junctions)
49
Early membrane function
40
Small outer capsid protein
soc
Core or scaffold:
gp22
ipI
ipII
ipIII
gp67
gp68
gp inh
Shell:
gp24
gp23
gp21
gp20
Unprocessed prohead
Assembling prohead
gro
gp31
gp40
gp21→T4PPase
Cleave
leave membrane
ipII
ipIII
Internal proteins
ipI
DNA end-binding function
2=64
4=50=64
Unknown completion functions
Processed prohead
Replicating DNA
ATP
DNA ligase
Resolve Holliday junctions
gp49
gp16+gp17
gp hoc
wac
gp soc
Outer capsid
Neck
gp13
gp14
Complete T4 head
Package DNA
expand
gp2=64
gp4=50=64
?
gp16
gp17
Partially filled head
wac
13
14
Whiskers
Neck
15
16
17
"Packasome" components
18
19
20
67=pip
68
21
22
23
24
hoc
inh=lip
Portal protein
Internal proteins
Prohead protease
Scaffold protein
Shell protein
Vertex protein
Surface protein
Inhibitor of protease
31
Early membrane function
30
DNA ligase

Figure 18-6 Assembly and maturation of the T4 prohead. Locations of genes that control these processes are shown on the circular genome map. Tail genes interspersed among the head genes are indicated on the inside of the circle. The figure is a modification of that in Black and Showe (49).

for gp24* (256). Those mutations in gene *23* that bypass the *24* requirement can also affect head length (152).

T4 head precursors (proheads) are assembled around a core or scaffold, consisting mainly of gp22 and gp21, and also containing gp67, gp68, and gp (inh = lip). After prohead assembly is complete, the scaffold proteins are cleaved to short peptides, followed by expansion and stabilization of the capsid shell (50) by DNA (177, 178).

The vertex from which assembly is initiated (at the bacterial membrane) contains the portal protein gp20. The portal consists of a ring of 12 copies of gp20. Its role in packaging and ejection of DNA is discussed in chapter 6. In all dsDNA tailed bacteriophages studied so far, the portal protein is of similar construction, and in the *Bacillus subtilis* bacteriophage ϕ29 (chapter 22) it has been shown to be a DNA translocating machine engaged in moving DNA into the head by a rotation mechanism (269, 373).

Functional T4 heads of altered length can be assembled that contain DNA shorter or longer than the normal complement (50, 231, 232, 276, 278, 289). Electron microscopy shows that there are three main head length classes, whereas widths are usually the same. Normal, elongated (giant), and isometric T4 heads are all 85 nm wide (104). The T4 variant containing 0.68 normal DNA length appears structurally isometric (T = 13, Q = 13). The head length classes fit Q numbers of 13 for isometric (0.68 DNA length) phages, 17 and 18 (0.85 and 0.90 DNA lengths) for intermediates, and 21 for normal DNA length phages (illustrated in figure 18-7). Fractional DNA lengths, measured by sucrose gradient sedimentation and by genetic methods (276, 289), or by gel electrophoresis (231, 232), fit the model that these DNA molecules fill the head volume. These shorter and longer head lengths are related to the basic mechanism determining T4 head length (231, 232, 267). The surface lattice design was determined directly from cryo-electron micrographs of isometric heads (176, 311) following earlier studies on phage T2 (54).

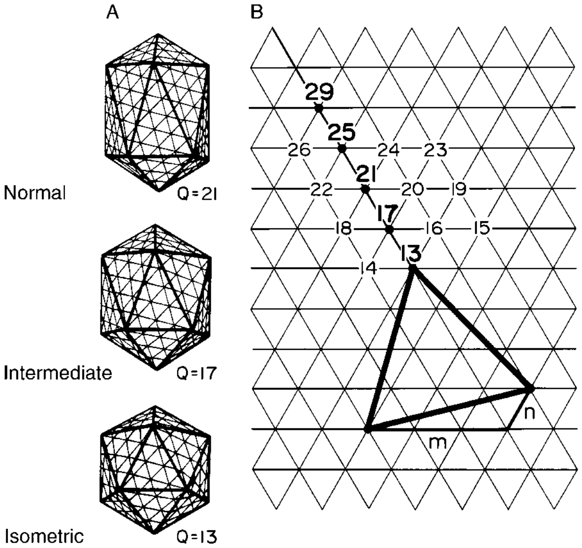


Figure 18-7 Surface lattice arrangements predicted for isometric, intermediate, and normal-length T4 heads. A: Representations of normal, intermediate, and isometric heads. The definitions of the lattice parameters are shown below. For the isometric phage, the surface lattice is defined completely by the triangulation number (T), given as $T(m,n) = m^2 + mn + n^2$; in the drawing $m = 3$, $n = 1$, and $T = 13$. For an icosahedron elongated along a 5-fold axis, another number, Q, is used to define the extent of elongation. B: The Q number can be any integer greater than T and is defined by the coordinates m' and n'. For the intermediate phage shown, $m' = 3$, $n' = 2$, and $Q = 17$. The normal-length head shown has $m' = 3$, $n' = 3$, and $Q = 21$; data from the laboratory of M. Rossmann et al. show that $Q = 20$ for their sample of T4 (113).

Head length is determined at a very early stage, prior to the formation of the unprocessed proheads. It is regulated primarily by protein–protein interactions between the major head components described below. Mutations in the genes for the major head protein (gene *23*), portal protein (gene *20*), scaffold protein (gene *22*), and vertex protein (gene *24*) can modulate head length (92, 93, 152, 353). Single amino acid changes in the gene *23* protein result in specifically altered head lengths. These mutations are located at 12 sites clustered in three locations in gene *23*. The total span of the three clusters is 155 bp, coding for approximately 52 amino acids (out of 521 amino acids in gp23). These sequences may represent segments of gp23* that interact with the other proteins that regulate head length, or with other gp23* subunits. Head length is also affected by variations in the relative intracellular concentrations of the head proteins (108), explaining virion size and density variations during the latent period (272).

Several models for regulation of head length (50) have been proposed (32, 33, 70, 195, 267, 322, 368). Of these, two models are currently favored, but both require experimental support and the real mechanism may combine aspects of both models. The Vernier model (33, 322) proposes that length is measured internally, like a mechanical vernier used in measuring devices. A possible example is two linear structures made of protein subunits of differing size. During copolymerization, the ends of the structures would only be in register after a certain number of subunits have been added, and polymerization would stop only when they are in register. Such a model was proposed as an explanation for the regulation of T4 head length—based on differing structural repeats between the elongating scaffold or core protein and the elongating coat protein—and has been refined more recently (33).

The cumulated strain or intrinsic curvature model proposes that as protein subunits assemble they are progressively distorted as a result of their bonding interactions which are due to the intrinsic curvature of the proteins. Ultimately, this distortion results in a switch of the entire structure into a lower energy conformation, where no further subunit addition is possible (195), or alternatively, the structure is directly regulated by the subunit curvature itself. Moody (267) has refined these arguments and defined two kinds of curvature, mean and Gaussian, that can account for the closed-shell assembly of spherical viruses. Clearly, in elongated T4 heads, additional size information must come from other sources, in this case the scaffold protein, since elongated heads have elongated scaffolds. In addition, Doherty's suppressor analyses (92, 93), mentioned above, support the influence of additional factors in head length determination. Moody's analysis is, however, an important theoretical contribution.

The observation that single amino acid changes in one head gene can change the head length distribution has important consequences for the interpretation of head lengths in T4-related phages. Phages that are related to T4 by sequence but which grow in different hosts and environments, may have isometric heads, like the marine cyanobacteriophage S-PM2 (148), or have longer heads than T4, like the vibriophage KVP40 (260). This implies that evolution has selected for plasticity, not rigidity, in bacteriophage capsid designs.

The T4 proheads expand during maturation and extensive protein structural arrangements take place. These result in large-scale movements of polypeptide chains from the surface to the inside and vice versa, as well as major shifts in inter-subunit connections (207). The expansion follows the initiation of DNA packaging (177, 178) and may "lock" the capsid into a stable, "toughened" structure. The early prohead lacks two accessory proteins of the complete T4 head, called Hoc and Soc, which are added after the head is expanded (50). Soc confers additional capsid stability, especially at low pH (175, 342). The Hoc protein lies at the center of a six-membered ring (capsomer) of gp23*, and there are six copies of Soc around each capsomer, except around the gp24* vertices (figure 18-3). The Hoc protein forms a mushroom-like projection that protrudes from the center of each capsomer (311) (figure 18-3). Osmotic shock resistance is influenced by the gp24* vertex protein (235, 303).

Studies of in vitro assembly of the T4 major head protein have so far provided limited insight into length regulation, but mutant studies have revealed important aspects of assembly in vivo. T4 mutants defective in head-assembly genes *20, 22,* and *40* accumulate large numbers of tubular assemblies of uncleaved gp23 called "polyheads." Mutants in genes *20* and *40* are defective in prohead initiation, and infected cells accumulate large quantities of unassembled gp23, which assemble into tubular polyheads late after infection (50). In gene *22* mutant-infected cells that express mutant gp22 internal scaffolding protein, polyheads are mostly of the same diameter, although the cylindrical pitch angles vary from tube to tube (267). However, in gene *22* mutants that completely lack the scaffold (core) protein, multilayered cylinders form around those tubes that assemble spontaneously, giving rise to tubes of different diameters.

In vitro, uncleaved monomeric gp23 protein (derived from dissociated polyheads) can be reversibly assembled into hexamers and polyheads, similar to assembly in other entropy-driven assembly systems such as flagella, microtubules, and tobacco mosaic virus (50). The information needed to form tubes is contained in the primary amino acid sequence of the head protein under favorable environmental conditions. Information for the assembly of hexamers, and for determination of the shell diameter, is also contained within the gp23 protein sequence.

Six trimeric "whiskers," each 53 nm long, are attached to the lower of the two knobs protruding from the portal head vertex (figure 18-3). The whiskers are made of α-helical, coiled-coil trimers of fibritin, the 486 amino acid product of the *wac* gene (394). The extended C-terminal β-annulus region initiates and stabilizes assembly of the trimeric whisker, and the N-terminal end binds to the neck of the head (52). These whiskers promote long tail fiber attachment, extension, and retraction (78, 435, 436), and serve as part of the environmental sensing system of the phage that maintains the long fibers in a retracted and protected configuration until needed. Figure 18-6 summarizes the assembly pathway of the T4 phage head and shows the map positions of the genes coding for capsid components.

Tail Structure and Assembly

Tails of *Myoviridae*, such as T4, have a remarkably constant structural design. Each tail has an intricate baseplate attached to a two-layered tube: an inner tubular part assembled from identical subunits of a roughly 20 kDa protein, arranged in 4.0 nm stacked disks, surrounded by a tubular contractile sheath assembled from a roughly 60 kDa protein with the same periodicity (53). In addition, tail fibers are usually attached to the baseplate. The baseplate and tail fibers are composed of many different parts that undergo coordinated conformational changes during infection, resulting in cell-surface binding, sheath contraction, membrane penetration, and DNA ejection.

T4 tail assembly is initiated by assembly of the complex baseplate (figure 18-3). T4 baseplates consist of a central component ("hub"), six outer "wedges," and six tail "pins" (81). Hub and wedge components assemble via independent pathways that subsequently converge to form complete baseplates.

The central hub of the baseplate remains poorly defined. Baseplate assembly initiates with formation of a multiprotein hub complex, which subsequently undergoes significant rearrangements before completion of assembly. The composition of the initial hub complex is unclear, but it at least contains gp29, which later adopts an extended form within the tail tube, and a complex of gp5 and gp27, which is used to penetrate the cell envelope during infection (187). (In addition, gp51, gp26, and gp28 may also be involved as assembly catalysts or structural parts.) The gp5/gp27 complex forms a membrane-puncturing needle that has lysozyme activity. When gp5 is incorporated into the baseplate, it is proteolytically cleaved (at Ser351–Ala352), and both cleavage parts remain in the baseplate. This cleavage is required to activate the lysozyme, which is used to break down the cellular peptidoglycan layer at the time of infection (189, 190, 292, 304). Near the C-terminal end of gp5 is an 11 nm long, 2.8 nm wide three-stranded β-helix that forms the needle, while the central lysozyme domain of gp5 lies around the baseplate-proximal part of the needle (187). A trimer of gp27 forms a hollow cylinder into which the N-terminal antiparallel β-barrel domain of gp5 is inserted, forming a thin, rigid device that facilitates penetration of the tail tube and viral DNA into the bacterial cell.

The six outer "wedges" form the bulk of the baseplate. Assembly initiates with the outer-edge proteins, and progresses toward the center. The earliest step in wedge assembly is the interaction of the tail pin proteins, gp10 and gp11, with gp7 and gp8. Other proteins that complete the wedge are added in the order gp6, gp53, and gp25 (81). Crowther et al. (84) determined the architecture of the hexagonal and star-shaped baseplates to about 3.0 nm resolution. From mutational analyses they deduced that three of the proteins studied (gp9, gp11, and gp12), account for 40% of the total mass of the baseplate, and they can all be added to the hexagonal structure after it is completed. Gp9 was determined to be the site of tail fiber binding since gp9 is needed for tail fiber attachment, is located near the site where tail fibers join to the baseplate, and antiserum directed at gp9 determinants blocks fiber attachment. Gp11 makes up the distal portion of the tail pin, and also supplies the gp12 binding site. Gp11 has been found to be immunologically related to gp10 (36) and may have evolved from it by gene duplication. The structure of gp11 has been determined to atomic resolution (236). Gp12 forms the six 35 nm short tail fibers, which implement irreversible binding to the host during infection (197). In wild-type phage and isolated tails, these fibers occupy a stored or folded position (187, 249). Six complete wedges assemble around a hub complex, and baseplate assembly is completed by addition of gp48 and gp54, providing the substrate for tail tube polymerization.

The inner tail tube is composed of 144 copies of a single protein, gp19, which are arranged in 24 stacked rings spaced 4.1 nm apart (81). The tube has an average diameter of 9.0 nm and an inner hole of 3.5 nm (268). The central hole in T4 tube baseplates (sheathless tails) is occupied by mass that is absent in tube baseplates partially emptied by treatment with guanidine HCl. This mass probably represents the gp29 "tape measure" protein that determines tail length (81, 101), similar to the gene H protein of phage λ. During assembly several copies of this protein present in the baseplate hub are thought to extend inside the growing tail tube and act as a template to limit tail length during tube assembly. It must also move out of the tube ahead of the DNA during ejection, demonstrating a remarkable series of activities for a single protein.

The contractile sheath, formed from 144 gp18 tail sheath subunits, is arranged nearly identically to gp19 subunits in the tail tube. All gp18 subunits are apparently in close contact with gp19 subunits of the tail tube since assembly of the sheath in an extended form is dependent on polymerization of the tail tube. The ring of gp18 nearest the head binds to the tail terminator, made of gp15 and gp3. Gp18 subunits in each ring are rotated by about 17° to the right with respect to those below, giving rise to the prominent right-handed helix seen on the surface (figure 18-3). Amos and Klug (11) portray each gp18 in the extended sheath as sloping downward from inner to outer radius, and this is also visible in the reconstructions of frozen hydrated tails (237). There is less downward slope in the model of Smith and Aebi (374). The sheath contracts during infection, due to a transition in which the axial repeat decreases from 4.1 nm to 1.5 nm and the twist angle changes from 17° to 32°. The dramatic change in the shape of the sheath upon contraction was shown to be related to relatively small changes in the overall conformation of gp18 (11, 196).

Long and Short Tail Fiber Structure and Assembly

The six long tail fibers of T4 are oriented for assembly onto the baseplate by the fibritin (wac) "whiskers" (436). Some other structure must serve this function in T2, since it lacks whiskers. Fibers about 3.0–4.0 nm thick and 150 nm long are made by the joining of two half-fibers at a kinked angle of about 150°. The half-fiber bound to the baseplate is constructed from three molecules of gp34 (71), and much of the fiber is likely composed of a three-stranded, β-helical structure. The half-fiber that binds to the cell surface is more complex, and is made of gp35 and three copies each of gp36 and gp37. The thin tip of gp37 contacts the cell surface with the C-terminal end (21).

From combined X-ray diffraction and electron microscopic studies, it has been proposed that the distal half-fiber is composed of a set of globular domains at specific regions along the half-fiber and that many of the polypeptide chains are in a cross-β conformation with face-to-face packing of both gp36 and gp37.

T4 fiber assembly requires the catalytic function of gp38 (436). However, in phage T2 and several T-even-like phages, a non-homologous gp38 participates with gp37 in host range determination and receptor recognition (156). This protein functions as an adhesin, or cell-surface-binding protein, similar to those found in the pili of bacterial pathogens.

Throughout the assembly process, host cell functions are required for proper construction of many parts of the virus. The GroEL protein chaperone and phage gp31 co-chaperonin are needed for head assembly. Tail fiber assembly also depends on the interaction of viral gp57A with the fibers, but host mutants of *E. coli* have been isolated that do not require gene *57A* function (436).

The short tail fibers (gp12) at the bottom of the baseplate have a 24 nm shaft, 3.8 nm in diameter at the N-terminus, and a C-terminal globular domain (249). The short tail fibers have two distinct conformations. As part of the hexagonal baseplate they exist in a compact, "stored" position with a bend near the center. When the hexagonal baseplate expands to the star-shaped conformation, the fibers extend to full length. The crystal structure of a heat- and protease-stable part of the fiber (residues 85–396 and associated residues 518–527) shows an N-terminal elongated domain, a striking right-hand triple β-helix domain (290–327) and a C-terminal, more globular domain (408, 409).

The short fibers attach to baseplate protein gp11, itself a trimer, each monomer of which has 218 residues folded into three domains (236). Both in the assembly pathway and in the structure, trimeric gp11 binds to trimeric gp10. The gp10 trimer binds to both trimeric gp9 and gp11. The C-terminus of gp11 lies at the gp10–gp11 interface, while the N-terminus binds the short fibers. Likewise, the C-terminus of gp9 binds to gp11, and the N-terminus of trimeric gp9 binds the trimeric long half-fiber gp34. Thus, when the long fiber binds reversibly to the cell surface, a signal would be transmitted via gp9 through gp10 and gp11 to the short fiber gp12, activating it for tight binding to the cell (236).

Assembly of both the long and short tail fibers requires participation of the viral-coded chaperone gp57A; in its absence gp34, gp37, and gp12 fail to trimerize and form intracellular, insoluble aggregates (60, 61). Unlike the GroEL/ES or the GroEL/gp31 chaperonins, which are relatively nonspecific regarding substrate proteins, gp57A appears to act more specifically by blocking aggregation of folding intermediates of gp12, gp34, and gp37, each of which contains complex three-stranded β-helices that require chaperone action.

Evolution

As mentioned earlier, the "essential" genes of the most closely related T-even phages are arranged in the same order, and their sequences are similar, but the genomes of different members of the family have different "nonessential" genes interspersed between these essential genes. These differences first became evident as insertion or substitution loops in electron micrographs of heteroduplex DNA prepared in vitro by annealing single strands of T2, T4, and T6 DNA (201) (figure 18-2), and have been confirmed by sequence comparisons in many cases (128, 226, 290, 396, 397). Although inactivating the nonessential genes has no consequences for phage growth in the laboratory, some are suspected to be important under different physiological conditions, and in the face of various restriction systems imposed by different hosts. The more distantly related T-even phages (2) have also rearranged the order of their essential genes.

Based on similarity or divergence of tail and head genes, the T4-related phages have been classified as T-even, pseudo-T-even and schizo-T-even phages (397). However, such classifications are arbitrary for any phages because horizontal gene transfer has resulted in extensive mosaics of phage genomes (233). For example, T4 and T6, which are closely related in the tail genes, differ in their dCTPase genes as much as the pseudo-T-even phages differ in their tail genes (S. Lousteau and G. Mosig, unpublished data). Although it has been postulated that the T4-related phages have experienced few exchanges with other phage groups (226, 396, 398), there is now ample evidence to the contrary (153, 233; chapter 21).

The view that "all the world's a phage" (155) is supported by patchy similarities of tail fiber genes of many phages (130, 145, 154, 156, 259, 339). Moreover, the T4 endonuclease VII (gp49) (18, 200), which cleaves recombinational intermediates, resembles an endonuclease of *Mycobacterium* phages L5 and D29 (155). The T4 DNA ligase (gp30) resembles the T7 DNA ligase (15). The sigma factor for late T4 transcription, gp55, resembles other sigma factors (140, 142). Gp17, part of the ATP-dependent terminase and DNA packaging complex, has regions of homology with ATP binding packaging proteins of other phages and of herpesviruses. The T4 lysozymes (gp e and gp5) resemble lysozymes of other phages (127, 292, 423). This list will undoubtedly become larger as more complete genomes are sequenced.

Exchanges between different phage genomes were first experimentally analyzed among the lambdoid phages, leading to the conclusion that entire modules of related genes had to be exchanged together (64). In retrospect, the strict concept of modules may be imposed on phages with site-specific recombination systems and few control regions, such as promoters. In contrast, the T4-related phages have a much higher potential for homologous recombination, and they have many independent promoters and overlapping transcription units. This allows lateral transfer via ectopic homologous recombination of much smaller units, that is, genes and gene segments. In fact, horizontal transfer of different genes to a site adjacent to the dCTPase gene in different T-even phages is apparently responsible for the large differences found between the dCTPase genes of different T-even phages (285, 290). We have proposed that this mechanism involves pairing of partially homologous sequences, join-cut-copy recombination and heteroduplex repair (figure 18-8), and that it generates multiple mutations from a single horizontal transfer. This mechanism is probably especially prevalent among the T4-related phages because of their high recombination potential. It can explain the differential variability of other phage genes as well, and throws doubt on interpretations of cladistic analyses.

For example, substitutions of foreign sequence into the tail fiber genes can account for the differences in tail fibers and host range of among various members of the family and for the remarkably rapid coevolution of viral and host genomes (156). The T4 terminase subunit gp17 has at its N-terminal segment a ssDNA binding domain that is lacking in terminases of other phages. However, shorter proteins can be synthesized from transcripts initiated at an internal promoter (119, 118). We have discussed the evolutionary relevance under "DNA Packaging." The lysozyme domain of T4 gp5 (discussed in the "Tail Structure and Assembly") is missing in gp5 of the related *Vibrio* phage KVP40 (E. Miller, personal communication). Whereas such differences could also be explained by deletions of gene segments, the DNA sequence divergence in adjacent regions suggests that these domains have been acquired by lateral transfer via a mechanism depicted in figure 18-8. In turn, the heterologies contribute to the species barriers between different members of the family by reducing production of viable recombinants.

Remarkably, T4, but so far no other member of this family, encodes several related endonucleases, either embedded into introns and responsible for homing of introns (24, 26, 27), or free-standing: *segA–F* (28, 363) and *mobA–mobF* (262, 400). These endonucleases can contribute to T4's ability to exclude alleles of T4-like phages, by making specific cuts in T4-like phages but not in T4 (28, 290). However, other nucleases must contribute to the mutual exclusion of T4-like phages since exclusion also occurs between phages that do not contain any of the endonucleases mentioned above (G. Mosig, unpublished data).

DNA of T4 and its relatives has a much higher A-T content than *E. coli* DNA, and there are different opinions as to how this difference is maintained. A comparison of the homologous T4 and *E. coli* Dam methylase genes (151) suggests

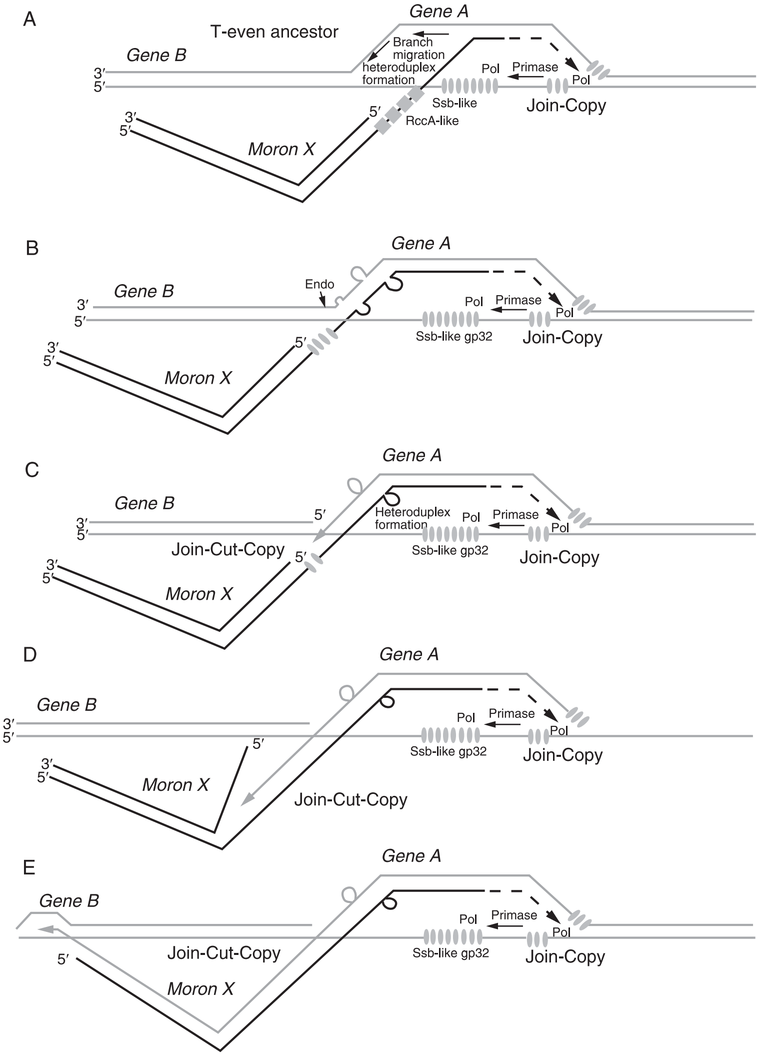

A
T-even ancestor
Gene A
Gene B
Branch migration
heteroduplex formation
Pol
Primase
Pol
Ssb-like
RecA-like
Join-Copy
Moron X
3′
5′
B
Gene A
Gene B
Endo
Pol
Primase
Pol
Ssb-like gp32
Join-Copy
Moron X
C
Gene A
Gene B
Heteroduplex formation
Pol
Primase
Pol
Ssb-like gp32
Join-Cut-Copy
Join-Copy
Moron X
D
Gene A
Gene B
Pol
Primase
Pol
Ssb-like gp32
Join-Copy
Moron X
Join-Cut-Copy
E
Gene A
Gene B
Pol
Primase
Pol
Ssb-like gp32
Join-Cut-Copy
Join-Copy
Moron X

Figure 18-8 Moron acquisition. A model to explain acquisition of a new gene or gene segment (moron) by an adaptation of join-cut-copy recombination (pathway III in figure 18-4). Very limited sequence homology between the foreign DNA and a resident gene A allows single-strand invasion by the fragment's end. Branch migration into a heterologous region generates heteroduplexes with multiple loops and mismatches, which can be repaired by endo VII (198, 295, 296). Initiation of replication from an endonuclease cut in the invaded strand will copy the moron into the ancestral T-even genome, if a short homology at the other end of the foreign DNA allows a similar invasive recombination. From (285). See thebacteriophages.org/frames_0180.htm for a color version of this figure.

that silent codon changes corresponding to T4 tRNA anticodons are rapidly selected. For example, CUG is the most frequently used *E. coli* leucine codon, and 13 of the 26 leucine codons are CUG. In contrast, none of the 29 T4 leucine codons is CUG, in keeping with the fact that *E. coli* leucyl tRNA is rapidly cleaved after T4 infection. In most other genes the AT-rich codons corresponding to T4 tRNAs predominate. The notable exception is the highly expressed gene *23*. This led Kunisawa et al. (224) to reject the above theory.

In any case, the differences in A-T content make alignments of T4 DNA sequences with homologous sequences of other organisms tenuous, but alignments of amino acid sequences can reveal evolutionary relationships. Both aerobic and anaerobic ribonucleotide reductases of T4 closely resemble the *E. coli* ribonucleotide reductase (372). In some cases, structural analyses have revealed similarities not detectable at the sequence levels. The first examples were T4 and goose lysozymes (254).

It now appears that the DNA replication proteins of T4 are more closely related to the replication proteins of eukaryotic viruses and their hosts than to those of *E. coli* (116, 411). Similarly, some recombination genes of T4 share sequence similarities with corresponding proteins of all life forms: the UvsX protein with RecA-like proteins (106), the gp46–47 complex with the SbcBC and Mre11–Rad50 complex (79, 80) and with the RecBC enzyme of *E. coli* (39). Based on such similarities it has been postulated that transfer of these genes from DNA viruses has shaped the evolution of eukaryotes.

Outlook

As suggested throughout this chapter, many complexities and redundancies of T4 are perhaps best understood in terms of its evolutionary history. Exchange of genes and acquisition of genes from other genomes (by lateral transfer) has been the raw material for natural selection under different environmental conditions. It is already well established that many genes are shared by rather different phages growing in different hosts and environments (155). Moreover, phages with contractile tail sheaths that resemble T4 have been found in many environments (148) (Robert V. Miller, personal communication). They contain many essential genes that are conserved in sequence, and often in relative locations, interspersed with different nonessential or auxiliary genes of unknown function. Sequencing of their complete genomes is bound to contribute to understanding the evolution of phages specifically, and of evolution in general. It should also reveal to what extent site-specific recombination, transposition, and ectopic homologous recombination contributed to lateral gene transfer in different phage genomes.

Two major lessons from T4, mentioned throughout this chapter, are: (i) The concept of protein machines, first developed to describe the exquisite structures formed during virion assembly, and their conformational changes during assembly and infection has been extended to the smaller, more fragile macromolecular assemblies that drive, among others, DNA replication, DNA precursor biosynthesis, recombination, repair, packaging and transcription. It remains to be seen to what extent the rules governing these protein nanomachines are universal for assemblies of defined stochiometry as compared with assemblies that change their functions by acquiring different parts. (ii) The processes that are driven by these protein machines are tightly interrelated, by physiological connections as well as by sharing of certain proteins. Thus, neither functioning nor evolution of these mechanisms and processes can be understood in isolation.

DNA sequences and deduced protein sequences can now be used to evaluate possible general principles. Three aspects learned from such studies are as follows: (i) Although T4 proteins share certain similarities with proteins of similar function in other systems, in most cases the similarities are patchy, and they are not apparent at the DNA level. (ii) Some T4 proteins, seemingly unrelated, share patches of homology, even at the DNA level. (iii) Many T4 genes produce more than one in-frame protein, probably because multiple forms are important in the proper assembly and function of the corresponding protein assemblies.

We have discussed elsewhere (277, 279) how the original strategy of T4 DNA replication together with the strategy of the transcriptional program provides a most powerful selection for efficient recombination. Recombination between partially homologous sequences, in turn, can initiate lateral gene transfer, associated with multiple sequence changes (279, 285, 290). It can also generate gene fusions and even proteins with novel functions. The homologies of T-even and eukaryotic replication proteins allowed general conclusions from exquisite biochemical, biophysical, and genetic analyses of T-even proteins. Short similarity of some T4 sequences with sequences of its host, other phages, plasmids, and eukaryotic genomes probably provides the substrate for "mix and match" evolution. The sharing of amino acid sequences among replication, transcription, and packaging components, and recombination between the corresponding DNA, might facilitate the rapid coevolution of these components and their integration into the complex T4 system.

Because the essential genes of T4 have been exquisitely characterized by many investigators, in terms of both regulation of their expression and functions of their products, the analysis of its annotated genome (262) will serve as a guide to interpretations of related genomes, which are now rapidly being sequenced (87) (J. Karam, personal communication; see http://phage.bioc.tulane.edu). The lessons learned from T4 alert us to the dangers of strictly computational methods, but encourage creative experimentation. In turn, these comparisons will undoubtedly refine our understanding of virus evolution specifically as well as of general evolutionary principles.

One of the more general aspects of T4 biology is the remarkable redundancy of pathways and proteins for fundamental processes, such as DNA replication and recombination. The importance of redundant pathways for global evolution of novel developmental circuits has been discussed (203). To maintain redundant pathways during evolution, it is important that they are based on different principles and subject to different pressures. The persistence of mechanistically different redundant pathways in T4 can be readily rationalized because they are best adapted to different stages of transcription, gene expression in general, and DNA packaging during the course of phage development (279, 287). Therefore, these redundancies are ready to bypass certain lesions in other important genes, and they confer tolerance to changes that facilitate evolution.

Acknowledgments

During the preparation of this article, our friend and co-author Gisela Mosig passed away. She will always be remembered for her intellectual acuity, and her warmth and caring as a mentor and colleague.

We thank past and present members of our laboratories for many stimulating discussions and comments, especially Nancy Canan for sequencing and Suzanne Lousteau for creative sequence analyses. G. M. thanks NSF MCB 9983568 for financial support, the Vanderbilt Ingram Cancer Center supported by NIH CA for facilities, and Dana Franklin for improving the manuscript. F. E. is grateful to Mari Gingery, Margaret Kowalczyk, and Jeanne Brady for creative assistance.

References

1. Abedon, S. T. 1994. Lysis and the interaction between free phages and infected cells, pp. 397–405. *In* J. D. Karam, J.W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L.W. Black, E. K. Spicer, E. Kutter, C. Carlson, and E. S. Miller (eds.) Molecular Biology of Bacteriophage T4. American Society for Microbiology, Washington, D.C.
2. Ackermann, H.W., and H. M. Krisch. 1997. A catalogue of T4-type bacteriophages. Arch. Virol. 142:2329–2345.
3. Adari, H. Y., K. Rose, K. R. Williams, W. H. Konigsberg, T. C. Lin, and E. K. Spicer. 1985. Cloning, nucleotide sequence, and overexpression of the bacteriophage T4 regA gene. Proc. Natl. Acad. Sci. USA. 82:1901–1905.
4. Adari, H.Y., and E. K. Spicer. 1986. Translational repression in vitro by the bacteriophage T4 regA protein. Proteins 1:116–124.
5. Adelman, K., E. N. Brody, and M. Buckle. 1998. Stimulation of bacteriophage T4 middle transcription by the T4 proteins MotA and AsiA occurs at two distinct steps in the transcription cycle. Proc. Natl. Acad. Sci. USA. 95:15247–15252.
6. Adelman, K., G. Orsini, A. Kolb, L. Graziani, and E. N. Brody. 1997. The interaction between the AsiA protein of bacteriophage T4 and the sigma70 subunit of *Escherichia coli* RNA polymerase. J. Biol. Chem. 272:27435–27443.
7. Alberts, B. 1998. The cell as a collection of protein machines: preparing the next generation of molecular biologists. Cell 92:291–294.
8. Alberts, B., and R. Miake-Lye. 1992. Unscrambling the puzzle of biological machines: the importance of the details. Cell 68:415–420.
9. Alberts, B. M. 1984. The DNA enzymology of protein machines. Cold Spring Harb. Symp. Quant. Biol. 49:1–12.
10. Alberts, B. M., and L. Frey. 1970. T4 bacteriophage gene 32: a structural protein in the replication and recombination of DNA. Nature 227:1313–1318.
11. Amos, L. A., and A. Klug. 1975. Three-dimensional image reconstructions of the contractile tail of T4 bacteriophage. J. Mol. Biol. 99:51–64.
12. Andrake, M. D., and J. D. Karam. 1991. Mutational analysis of the mRNA operator for T4 DNA polymerase. Genetics 128:203–213.
13. Ang, D., A. Richardson, M. P. Mayer, F. Keppel, H. Krisch, and C. Georgopoulos. 2001. Pseudo-T-even bacteriophage RB49 encodes CocO, a cochaperonin for GroEL, which can substitute for *Escherichia coli*'s GroES and bacteriophage T4's Gp31. J. Biol. Chem. 276:8720–8726.
14. Appasani, K., D. S. Thaler, and E. B. Goldberg. 1999. Bacteriophage T4 gp2 interferes with cell viability and with bacteriophage lambda Red recombination. J. Bacteriol. 181:1352–1355.
15. Armstrong, J., R. S. Brown, and A. Tsugita. 1983. Primary structure and genetic organization of phage T4 DNA ligase. Nucleic Acids Res. 11:7145–7156.
16. Baker, T. S., N. H. Olson, and S. D. Fuller. 1999. Adding the third dimension to virus life cycles: three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs. Microbiol. Mol. Biol. Rev. 63:862–922.
17. Barcus, V. A., A. J. Titheradge, and N. E. Murray. 1995. The diversity of alleles at the hsd locus in natural populations of *Escherichia coli*. Genetics 140:1187–1197.
18. Barth, K. A., D. Powell, M. Trupin, and G. Mosig. 1988. Regulation of two nested proteins from gene 49 (recombination endonuclease VII) and of a lambda RexA-like protein of bacteriophage T4. Genetics 120:329–343.
19. Bautz, F. A., and E. K. Bautz. 1967. Transformation in phage T4: minimal recognition length between donor and recipient DNA. Genetics 57:887–895.
20. Bebenek, A., G. T. Carver, H. K. Dressman, F. A. Kadyrov, J. K. Haseman, V. Petrov, W. H. Konigsberg, J. D. Karam, and J. W. Drake. 2002. Dissecting the fidelity of bacteriophage RB69 DNA polymerase: site-specific modulation of fidelity by polymerase accessory proteins. Genetics 162:1003–1018.
21. Beckendorf, S. K. 1973. Structure of the distal half of the bacteriophage T4 tail fiber. J. Mol. Biol. 73:37–53.
22. Beernink, H. T., and S. W. Morrical. 1999. RMPs: recombination/replication mediator proteins. Trends Biochem. Sci. 24:385–389.
23. Beernink, H. T., and S. W. Morrical. 1998. The uvsY recombination protein of bacteriophage T4 forms hexamers in the presence and absence of single-stranded DNA. Biochemistry 37:5673–5681.
24. Belfort, M., V. Derbyshire, M. Parker, B. Cousineau, and A. M. Lambowitz. 2002. Mobile introns: pathways and proteins, pp. 761–783. *In* N. L. Craig, R. Craigie, M. Gellert, and A. M. Lambowitz (eds.) Mobile DNA II. ASM Press, Washington, D.C.
25. Belfort, M., J. Pedersen-Lane, K. Ehrenman, F. K. Chu, G. F. Maley, F. Maley, D. S. McPheeters, and L. Gold. 1986. RNA splicing and in vivo expression of the intron-containing td gene of bacteriophage T4. Gene 41:93–102.
26. Belfort, M., and P. S. Perlman. 1995. Mechanisms of intron mobility. J. Biol. Chem. 270:30237–30240.
27. Belfort, M., and R. J. Roberts. 1997. Homing endonucleases: keeping the house in order. Nucleic Acids Res. 25: 3379–3388.
28. Belle, A., M. Landthaler, and D. A. Shub. 2002. Intronless homing: site-specific endonuclease SegF of bacteriophage T4 mediates localized marker exclusion analogous to homing endonucleases of group I introns. Genes Dev. 16:351–362.
29. Benkovic, S. J., A. M. Valentine, and F. Salinas. 2001. Replisome-mediated DNA replication. Annu. Rev. Biochem. 70:181–208.
30. Benzer, S. 1957. The elementary units of heredity, pp. 70–93. *In* W. D. McElroy and B. Glass (eds.) The

Chemical Basis of Heredity. The Johns Hopkins Press, Baltimore, Md.

31. Benzer, S., and S. P. Champe. 1962. A change from nonsense to sense in the genetic code. Proc. Natl. Acad. Sci. USA 48:1114–1121.
32. Berger, B., G. W. Hoest, J. R. Paulson, and P. W. Shor. 1999. On the structure of the scaffolding core of bacteriophage T4. J. Comput. Biol. 6:1–12.
33. Berger, B., and P. W. Shor. 1998. on the structure of the scaffolding core of bacteriophage T4 and its role in head length determination. J. Struct. Biol. 121:285–294.
34. Berger, H., and A. W. Kozinski. 1969. Suppression of T4D ligase mutations by rIIa and rIIb mutations. Proc. Natl. Acad. Sci. USA 64:897–904.
35. Berger, H., A. J. Warren, and K. E. Fry. 1969. Variations in genetic recombination due to amber mutations in T4D bacteriophage. J. Virol. 3:171–175.
36. Berget, P. B., and J. King. 1978. Isolation and characterization of precursors in T4 baseplate assembly. The complex of gene 10 and gene 11 products. J. Mol. Biol. 124: 469–486.
37. Bernstein, C., and S. S. Wallace. 1983. DNA repair, pp. 138–151. *In* C. K. Mathews, E. Kutter, G. Mosig, and P. B. Berget (eds.) Bacteriophage T4. American Society for Microbiology, Washington, D.C.
38. Bernstein, H. 1968. Repair and recombination in phage T4. I. Genes affecting recombination. Cold Spring Harb. Symp. Quant. Biol. 33:325–331.
39. Bernstein, H., and C. Bernstein. 1989. Bacteriophage T4 genetic homologies with bacteria and eucaryotes. J. Bacteriol. 171:2265–2270.
40. Bhagwat, M., and N. G. Nossal. 2001. Bacteriophage T4 RNase H removes both RNA primers and adjacent DNA from the 5′ end of lagging strand fragments. J. Biol. Chem. 276:28516–28524.
41. Bhattacharyya, S. P., and V. B. Rao. 1993. A novel terminase activity associated with the DNA packaging protein gp17 of bacteriophage T4. Virology 196:34–44.
42. Bhattacharyya, S. P., and V. B. Rao. 1994. Structural analysis of DNA cleaved in vivo by bacteriophage T4 terminase. Gene 146:67–72.
43. Birkenbihl, R. P., and B. Kemper. 1998. Endonuclease VII has two DNA-binding sites each composed from one N- and one C-terminus provided by different subunits of the protein dimer. EMBO J. 17:4527–4534.
44. Birkenbihl, R. P., and B. Kemper. 1998. Localization and characterization of the dimerization domain of holliday structure resolving endonuclease VII of phage T4. J. Mol. Biol. 280:73–83.
45. Birkenkamp, K., and B. Kemper. 1996. Bacteriophage T4 strand transfer protein UvsX tolerates symmetric and asymmetric heterologies in short double-stranded oligonucleotides. J. Mol. Biol. 259:622–631.
46. Birkenkamp, K., and B. Kemper. 1995. In vitro processing of heteroduplex loops and mismatches by endonuclease VII. DNA Res. 2:9–14.
47. Birkenkamp-Demtroder, K., S. Golz, and B. Kemper. 1997. Inhibition of Holliday structure resolving endonuclease VII of bacteriophage T4 by recombination enzymes UvsX and UvsY. J. Mol. Biol. 267:150–162.
48. Black, L. W. 1995. DNA packaging and cutting by phage terminases: control in phage T4 by a synaptic mechanism. Bioessays 17:1025–1030.
49. Black, L. W., and M. K. Showe. 1983. Morphogenesis of the T4 head, pp. 219–245. *In* C. K. Mathews, E. Kutter, G. Mosig, and P. B. Berget (eds.) Bacteriophage T4. American Society for Microbiology, Washington, D.C.
50. Black, L. W., M. K. Showe, and A. C. Steven. 1994. Morphogenesis of the T4 head, pp. 218–258. *In* J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, C. Carlson, and E. S. Miller (eds.) Molecular Biology of Bacteriophage T4. American Society for Microbiology, Washington, D.C.
51. Bleuit, J. S., H. Xu, Y. Ma, T. Wang, J. Liu, and S. W. Morrical. 2001. Mediator proteins orchestrate enzyme–ssDNA assembly during T4 recombination-dependent DNA replication and repair. Proc. Natl. Acad. Sci. USA 98:8298–8305.
52. Boudko, S. P., Y. Y. Londer, A. V. Letarov, N. V. Sernova, J. Engel, and V. V. Mesyanzhinov. 2002. Domain organization, folding and stability of bacteriophage T4 fibritin, a segmented coiled-coil protein. Eur. J. Biochem. 269: 833–841.
53. Bradley, D. E. 1967. Ultrastructure of bacteriophage and bacteriocins. Bacteriol. Rev. 31:230–314.
54. Branton, D., and A. Klug. 1975. Capsid geometry of bacteriophage T2: a freeze-etching study. J. Mol. Biol. 92: 559–565.
55. Brenner, S., F. Jacob, and M. Meselson. 1961. Unstable intermediate carrying information from genes to ribosomes for protein synthesis. Nature 190:41–46.
56. Broida, J., and J. Abelson. 1985. Sequence organization and control of transcription in the bacteriophage T4 tRNA region. J. Mol. Biol. 185:545–563.
57. Broker, T. R. 1973. An electron microscopic analysis of pathways for bacteriophage T4 DNA recombination. J. Mol. Biol. 81:1–16.
58. Broker, T. R., and A. H. Doermann. 1975. Molecular and genetic recombination of bacteriophage T4. Annu. Rev. Genet. 9:213–244.
59. Broker, T. R., and I. R. Lehman. 1971. Branched DNA molecules: intermediates in T4 recombination. J. Mol. Biol. 60:131–149.
60. Burda, M. R., I. Hindennach, and S. Miller. 2000. Stability of bacteriophage T4 short tail fiber. Biol. Chem. 381: 255–258.
61. Burda, M. R., and S. Miller. 1999. Folding of coliphage T4 short tail fiber in vitro. Analysing the role of a bacteriophage-encoded chaperone. Eur. J. Biochem. 265:771–778.
62. Busby, S., and R. H. Ebright. 1999. Transcription activation by catabolite activator protein (CAP). J. Mol. Biol. 293:199–213.
63. Calendar, R., S. Yu, H. Myung, V. Barreiro, R. Odegrip, K. Carlson, L. Davenport, G. Mosig, G. E. Christie, and E. Hagga°rd-Ljundquist. 1998. The lysogenic conversion genes of coliphage P2 have unusually high AT content, pp. 241–252. *In* M. Syvanen (ed.) Horizontal Gene Transfer. Chapman and Hall, London.
64. Campbell, A., and D. Botstein. 1983. Evolution of the lambdoid phages, pp. 365–380. *In* R. Hendrix, J. Roberts, F. Stahl, and R. Weisberg (eds.) Lambda II. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

65. Carles-Kinch, K., J. W. George, and K. N. Kreuzer. 1997. Bacteriophage T4 UvsW protein is a helicase involved in recombination, repair and the regulation of DNA replication origins. EMBO J. 16:4142–4151.
66. Carlson, K., and L. D. Kosturko. 1998. Endonuclease II of coliphage T4: a recombinase disguised as a restriction endonuclease? Mol. Microbiol. 27:671–676.
67. Carlson, K., L. D. Kosturko, and A. C. Nystrom. 1996. Short-range and long-range context effects on coliphage T4 endonuclease II-dependent restriction. J. Bacteriol. 178: 6419–6426.
68. Carlson, K., E. A. Raleigh, and S. Hattman. 1994. Restriction and modification, pp. 369–381. *In* J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, C. Carlson, and E. S. Miller (eds.) Molecular Biology of Bacteriophage T4. American Society for Microbiology, Washington, D.C.
69. Carpousis, A. J., and H. M. Krisch. 1994. mRNA processing and degradation, pp. 176–181. *In* J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, C. Carlson, and E. S. Miller (eds.) Molecular Biology of Bacteriophage T4. American Society for Microbiology, Washington, D.C.
70. Caspar, D. L. 1980. Movement and self-control in protein assemblies. Quasi-equivalence revisited. Biophys. J. 32: 103–138.
71. Cerritelli, M. E., J. S. Wall, M. N. Simon, J. F. Conway, and A. C. Steven. 1996. Stoichiometry and domainal organization of the long tail-fiber of bacteriophage T4: a hinged viral adhesin. J. Mol. Biol. 260:767–780.
72. Chace, K. V., and D. H. Hall. 1975. Characterization of new regulatory mutants of bacteriophage T4. II. New class of mutants. J. Virol. 15:929–945.
73. Chan, V. L., and K. Ebisuzaki. 1973. Intergenic suppression of amber polynucleotide ligase mutation in bacteriophage T4. II. Virology 53:60–74.
74. Chastain, P. D., 2nd, A. M. Makhov, N. G. Nossal, and J. D. Griffith. 2000. Analysis of the Okazaki fragment distributions along single long DNAs replicated by the bacteriophage T4 proteins. Mol. Cell 6:803–814.
75. Chauhan, A. K., and D. Apirion. 1991. The *rne* gene is the structural gene for the processing endoribonuclease RNase E of *Escherichia coli*. Mol. Gen. Genet. 228: 49–54.
76. Clyman, J., S. Quirk, and M. Belfort. 1994. Mobile introns in the T-even phages, pp. 83–88. *In* J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, C. Carlson, and E. S. Miller (eds.) Molecular Biology of Bacteriophage T4. American Society for Microbiology, Washington, D.C.
77. Colland, F., G. Orsini, E. N. Brody, H. Buc, and A. Kolb. 1998. The bacteriophage T4 AsiA protein: a molecular switch for sigma 70-dependent promoters. Mol. Microbiol. 27: 819–829.
78. Conley, M. P., and W. B. Wood. 1975. Bacteriophage T4 whiskers: a rudimentary environment-sensing device. Proc. Natl. Acad. Sci. USA 72:3701–3705.
79. Connelly, J. C., L. A. Kirkham, and D. R. Leach. 1998. The SbcCD nuclease of *Escherichia coli* is a structural maintenance of chromosomes (SMC) family protein that cleaves hairpin DNA. Proc. Natl. Acad. Sci. USA 95:7969–7974.
80. Connelly, J. C., and D. R. Leach. 2002. Tethering on the brink: the evolutionarily conserved Mre11–Rad50 complex. Trends Biochem. Sci. 27:410–418.
81. Coombs, D. H., and F. Arisaka. 1994. T4 tail structure and function, pp. 259–281. *In* J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, C. Carlson, and E. S. Miller (eds.) Molecular Biology of Bacteriophage T4. American Society for Microbiology, Washington, D.C.
82. Cox, M. M., M. F. Goodman, K. N. Kreuzer, D. J. Sherratt, S. J. Sandler, and K. J. Marians. 2000. The importance of repairing stalled replication forks. Nature 404:37–41.
83. Crick, F. H., L. Barnett, S. Brenner, and R. J. Watts-Tobin. 1961. General nature of the genetic code for proteins. Nature 192:1227–1232.
84. Crowther, R. A., E. V. Lenk, Y. Kikuchi, and J. King. 1977. Molecular reorganization in the hexagon to star transition of the baseplate of bacteriophage T4. J. Mol. Biol. 116: 489–523.
85. Dannenberg, R., and G. Mosig. 1983. Early intermediates in bacteriophage T4 DNA replication and recombination. J. Virol. 45:813–831.
86. Dannenberg, R., and G. Mosig. 1981. Semiconservative DNA replication is initiated at a single site in recombination-deficient gene 32 mutants of bacteriophage T4. J. Virol. 40:890–900.
87. Desplats, C., C. Dez, F. Tetart, H. Eleaume, and H. M. Krisch. 2002. Snapshot of the genome of the pseudo-T-even bacteriophage RB49. J. Bacteriol. 184:2789–2804.
88. Deutscher, M. P. 1993. Promiscuous exoribonucleases of *Escherichia coli*. J. Bacteriol. 175:4577–4583.
89. Dewey, M. J., and F. R. Frankel. 1975. Two suppressor loci for gene 49 mutations of bacteriophage T4. I. Genetic properties and DNA synthesis. Virology 68:387–401.
90. Dodson, M. L., and R. S. Lloyds. 1994. Structure-function analysis of T4 endonuclease V, pp. 318–321. *In* J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, C. Carlson, and E. S. Miller (eds.) Molecular Biology of Bacteriophage T4. American Society for Microbiology, Washington, D.C.
91. Doermann, A. H., and L. Boehner. 1963. An experimental analysis of bacteriophage T4 heterozygotes. I. mottled plaques from crosses involving six Rii loci. Virology 21:551–567.
92. Doherty, D. H. 1982. Genetic studies on capsid-length determination in bacteriophage T4. I. Isolation and partial characterization of second-site revertants of a gene *23* mutation affecting capsid length. J. Virol. 43:641–654.
93. Doherty, D. H. 1982. Genetic studies on capsid-length determination in bacteriophage T4. II. Genetic evidence that specific protein–protein interactions are involved. J. Virol. 43:655–663.
94. Drake, J. W. 1967. The length of the homologous pairing region for genetic recombination in bacteriophage T4. Proc. Natl. Acad. Sci. USA 58:962–966.
95. Drake, J. W., B. Charlesworth, D. Charlesworth, and J. F. Crow. 1998. Rates of spontaneous mutation. Genetics 148:1667–1686.
96. Drake, J. W., and K. N. Kreuzer. 1994. DNA transactions in T4-infected *Escherichia coli*, pp. 11–13. *In* J. D. Karam,

J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, C. Carlson, and E. S. Miller (eds.) Molecular Biology of Bacteriophage T4. American Society for Microbiology, Washington, D.C.

97. Drake, J. W., and L. S. Ripley. 1994. Mutagenesis, pp. 98–124. *In* J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, C. Carlson, and E. S. Miller (eds.) Molecular Biology of Bacteriophage T4. American Society for Microbiology, Washington, D.C.
98. Drivdahl, R. H., and E. M. Kutter. 1990. Inhibition of transcription of cytosine-containing DNA in vitro by the *alc* gene product of bacteriophage T4. J. Bacteriol. 172: 2716–2727.
99. Du, Z., D. P. Giedroc, and D. W. Hoffman. 1996. Structure of the autoregulatory pseudoknot within the gene *32* messenger RNA of bacteriophages T2 and T6: a model for a possible family of structurally related RNA pseudoknots. Biochemistry 35:4187–4198.
100. Duckworth, D. H. 1970. Biological activity of bacteriophage ghosts and "take-over" of host functions by bacteriophage. Bacteriol. Rev. 34:344–363.
101. Duda, R. L., M. Gingery, and F. A. Eiserling. 1986. Potential length determiner and DNA injection protein is extruded from bacteriophage T4 tail tubes in vitro. Virology 151:296–314.
102. Dudas, K. C., and K. N. Kreuzer. 2001. UvsW protein regulates bacteriophage T4 origin-dependent replication by unwinding R-loops. Mol. Cell Biol. 21: 2706–2715.
103. Dulbecco, R. 1949. Reactivation of ultraviolet-inactivated bacteriophage by visible light. Nature 163:949.
104. Earnshaw, W. C., J. King, S. C. Harrison, and F. A. Eiserling. 1978. The structural organization of DNA packaged within the heads of T4 wild-type, isometric and giant bacteriophages. Cell 14:559–568.
105. Edgar, R. S., and W. B. Wood. 1966. Morphogenesis of bacteriophage T4 in extracts of mutant-infected cells. Proc. Natl. Acad. Sci. USA 55:498–505.
106. Egelman, E. H. 2001. Does a stretched DNA structure dictate the helical geometry of RecA-like filaments? J. Mol. Biol. 309:539–542.
107. Eiserling, F. A., and L. W. Black. 1994. Pathways in T4 morphogenesis, pp. 209–217. *In* J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, C. Carlson, and E. S. Miller (eds.) Molecular Biology of Bacteriophage T4. American Society for Microbiology, Washington, D.C.
108. Eiserling, F. A., J. Corso, S. Feng, and R. H. Epstein. 1984. Intracellular morphogenesis of bacteriophage T4. II. Head morphogenesis. Virology 137:95–101.
109. Ellis, E. L., and M. Delbrück. 1939. The growth of bacteriophage. J. Gen. Physiol. 22:365–384.
110. Epstein, R. H., A. Bolle, C. Steinberg, E. Kellenberger, E. Boy de la Tour, R. Chevalley, R. Edgar, M. Susman, C. Denhardt, and I. Lielausis. 1964. Physiological studies of conditional lethal mutants of bacteriophage T4D. Cold Spring Harbor Symp. Quant. Biol. 28:375–392.
111. Finnin, M. S., M. P. Cicero, C. Davies, S. J. Porter, S. W. White, and K. N. Kreuzer. 1997. The activation domain of the MotA transcription factor from bacteriophage T4. EMBO J. 16:1992–2003.
112. Flemming, M., B. Deumling, and B. Kemper. 1993. Function of gene *49* of bacteriophage T4.III. Isolation of Holliday structures from very fast-sedimenting DNA. Virology 196:910–913.
113. Fokine, A., P. R. Chipman, P. G. Leiman, V. V. Mesyanzhinov, V. B. Rao, and M. G. Rossmann. 2004. Molecular architecture of the prolate head of bacteriophage T4. Proc. Natl. Acad. Sci. USA 101:6003–6008.
114. Formosa, T., and B. M. Alberts. 1986. DNA synthesis dependent on genetic recombination: characterization of a reaction catalyzed by purified bacteriophage T4 proteins. Cell 47:793–806.
115. Formosa, T., R. L. Burke, and B. M. Alberts. 1983. Affinity purification of bacteriophage T4 proteins essential for DNA replication and genetic recombination. Proc. Natl. Acad. Sci. USA 80:2442–2446.
116. Forterre, P. 2002. The origin of DNA genomes and DNA replication proteins. Curr. Opin. Microbiol. 5:525–532.
117. Frankel, F. R. 1966. Studies on the nature of replicating DNA in T4-infected *Escherichia coli*. J. Mol. Biol. 18:127–143.
118. Franklin, J. L., D. Haseltine, L. Davenport, and G. Mosig. 1998. The largest (70 kDa) product of the bacteriophage T4 DNA terminase gene *17* binds to single-stranded DNA segments and digests them towards junctions with double-stranded DNA. J. Mol. Biol. 277:541–557.
119. Franklin, J. L., and G. Mosig. 1996. Expression of the bacteriophage T4 DNA terminase genes *16* and *17* yields multiple proteins. Gene 177:179–189.
120. Franklin, M. C., J. Wang, and T. A. Steitz. 2001. Structure of the replicating complex of a pol alpha family DNA polymerase. Cell 105:657–667.
121. Frazier, M. W. 1989. PhD thesis, Vanderbilt University.
122. Frazier, M. W., and G. Mosig. 1990. The bacteriophage T4 gene *mrh* whose product inhibits late T4 gene expression in an *Escherichia coli* rpoH (sigma 32) mutant. Gene 88:7–14.
123. Fu, T. J., B. Kemper, and N. C. Seeman. 1994. Cleavage of double-crossover molecules by T4 endonuclease VII. Biochemistry 33:3896–3905.
124. Furukawa, H., T. Kuroiwa, and S. Mizushima. 1983. DNA injection during bacteriophage T4 infection of *Escherichia coli*. J. Bacteriol. 154:938–945.
125. Furukawa, H., H. Yamada, and S. Mizushima. 1979. Interaction of bacteriophage T4 with reconstituted cell envelopes of *Escherichia coli* K-12. J. Bacteriol. 140: 1071–1080.
126. Gaal, T., W. Ross, E. E. Blatter, H. Tang, X. Jia, V. V. Krishnan, N. Assa-Munt, R. H. Ebright, and R. L. Gourse. 1996. DNA-binding determinants of the alpha subunit of RNA polymerase: novel DNA-binding domain architecture. Genes Dev. 10:16–26.
127. Garvey, K. J., M. S. Saedi, and J. Ito. 1986. Nucleotide sequence of *Bacillus* phage phi 29 genes *14* and *15*: homology of gene *15* with other phage lysozymes. Nucleic Acids Res. 14:10001–10008.
128. Gary, T. P., N. E. Colowick, and G. Mosig. 1998. A species barrier between bacteriophages T2 and T4: exclusion,

join-copy and join-cut-copy recombination and mutagenesis in the dCTPase genes. Genetics 148:1461–1473.

129. Geiduschek, E. P. 1995. Connecting a viral DNA replication apparatus with gene Expression Semin. Virol. 6:25–33.

130. George, D. G., L. S. Yeh, and W. C. Barker. 1983. Unexpected relationships between bacteriophage lambda hypothetical proteins and bacteriophage T4 tail-fiber proteins. Biochem. Biophys. Res. Commun. 115:1061–1068.

131. George, J. W., B. A. Stohr, D. J. Tomso, and K. N. Kreuzer. 2001. The tight linkage between DNA replication and double-strand break repair in bacteriophage T4. Proc. Natl. Acad. Sci. USA 98:8290–8297.

132. Georgiou, T., Y. N. Yu, S. Ekunwe, M. J. Buttner, A. Zuurmond, B. Kraal, C. Kleanthous, and L. Snyder. 1998. Specific peptide-activated proteolytic cleavage of *Escherichia coli* elongation factor Tu. Proc. Natl. Acad. Sci. USA 95:2891–2895.

133. Gerald, W. L., and J. D. Karam. 1984. Expression of a DNA replication gene cluster in bacteriophage T4: genetic linkage and the control of gene product interactions. Genetics 107:537–549.

134. Gesteland, R. F., and J. F. Atkins. 1996. Recoding: dynamic reprogramming of translation. Annu. Rev. Biochem. 65:741–768.

135. Goldberg, E., L. Grinius, and L. Letellier. 1994. Recognition, attachment, and injection, pp. 347–356. *In* J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, C. Carlson, and E. S. Miller (eds.) Molecular Biology of Bacteriophage T4. American Society for Microbiology, Washington, D.C.

136. Goldberg, E. B. 1966. The amount of DNA between genetic markers in phage T4. Proc. Natl. Acad. Sci. USA 56:1457–1463.

137. Golz, S., A. Christoph, K. Birkenkamp-Demtroder, and B. Kemper. 1997. Identification of amino acids of endonuclease VII essential for binding and cleavage of cruciform DNA. Eur. J. Biochem. 245:573–580.

138. Golz, S., and B. Kemper. 1999. Association of Holliday-structure resolving endonuclease VII with gp20 from the packaging machine of phage T4. J. Mol. Biol. 285: 1131–1144.

139. Gordon, J., T. K. Sengupta, C. A. Phillips, S. M. O'Malley, K. R. Williams, and E. K. Spicer. 1999. Identification of the RNA binding domain of T4 RegA protein by structure-based mutagenesis. J. Biol. Chem. 274:32265–32273.

140. Gram, H., and W. Ruger. 1985. Genes *55*, *α gt*, *47* and *46* of bacteriophage T4: the genomic organization as deduced by sequence analysis. EMBO J. 4:257–264.

141. Grebenshchikova, S. M., L. A. Plugina, and V. P. Shcherbakov. 1994. The role of T4-bacteriophage endonuclease VII in correction of mismatched regions. Genetika 30: 622–626.

142. Gribskov, M., and R. R. Burgess. 1986. Sigma factors from *E. coli*, *B. subtilis*, phage SP01, and phage T4 are homologous proteins. Nucleic Acids Res. 14:6745–6763.

143. Gruidl, M. E., T. C. Chen, S. Gargano, A. Storlazzi, A. Cascino, and G. Mosig. 1991. Two bacteriophage T4 base plate genes (*25* and *26*) and the DNA repair gene *uvsY* belong to spatially and temporally overlapping transcription units. Virology 184:359–369.

144. Gurevitz, M., and D. Apirion. 1985. The ribonuclease-III-processing site near the 5′ end of an RNA precursor of bacteriophage T4 and its effect on termination. Eur. J. Biochem. 147:581–586.

145. Haggard-Ljungquist, E., C. Halling, and R. Calendar. 1992. DNA sequences of the tail fiber genes of bacteriophage P2: evidence for horizontal transfer of tail fiber genes among unrelated bacteriophages. J. Bacteriol. 174:1462–1477.

146. Hahn, S., U. Kruse, and W. Ruger. 1986. The region of phage T4 genes *34*, *33* and *59*: primary structures and organization on the genome. Nucleic Acids Res. 14:9311–9327.

147. Halpern, M. E., T. Mattson, and A. W. Kozinski. 1979. Origins of phage T4 DNA replication as revealed by hybridization to cloned genes. Proc. Natl. Acad. Sci. USA 76:6137–6141.

148. Hambly, E., F. Tetart, C. Desplats, W. H. Wilson, H. M. Krisch, and N. H. Mann. 2001. A conserved genetic module that encodes the major virion components in both the coliphage T4 and the marine cyanophage S-PM2. Proc. Natl. Acad. Sci. USA 98:11411–11416.

149. Harm, W. 1963. Mutants of phage T4 with increased sensitivity to ultraviolet. Virology 19:66–71.

150. Harvey, A., R. Vaiskunaite, and G. Mosig. 2002. Binding of an initiator protein RepEB to iterons upstream of an origin E primer promoter allows priming by a transcript of bacteriophage T4 DNA replication. Manuscript in preparation.

151. Hattman, S., J. Wilkinson, D. Swinton, S. Schlagman, P. M. Macdonald, and G. Mosig. 1985. Common evolutionary origin of the phage T4 dam and host *Escherichia coli* dam DNA-adenine methyltransferase genes. J. Bacteriol. 164: 932–937.

152. Haynes, J. A., and F. A. Eiserling. 1996. Modulation of bacteriophage T4 capsid size. Virology 221:67–77.

153. Hendrix, R. W. 1999. Evolution: the long evolutionary reach of viruses. Curr. Biol. 9:R914–R917.

154. Hendrix, R. W., and R. L. Duda. 1992. Bacteriophage lambda PaPa: not the mother of all lambda phages. Science 258:1145–1148.

155. Hendrix, R. W., M. C. Smith, R. N. Burns, M. E. Ford, and G. F. Hatfull. 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. Proc. Natl. Acad. Sci. USA 96:2192–2197.

156. Henning, U., and S. Hashemol-Hosseini. 1994. Receptor recognition by T-even-type coliphages, pp. 291–298. *In* J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, C. Carlson, and E. S. Miller (eds.) Molecular Biology of Bacteriophage T4. American Society for Microbiology, Washington, D.C.

157. Hercules, K., and J. S. Wiberg. 1971. Specific suppression of mutations in genes *46* and *47* by *das*, a new class of mutations in bacteriophage T4D. J. Virol. 8:603–612.

158. Herendeen, D. R., G. A. Kassavetis, J. Barry, B. M. Alberts, and E. P. Geiduschek. 1989. Enhancement of bacteriophage T4 late transcription by components of the T4 DNA replication apparatus. Science 245:952–958.

159. Herendeen, D. R., G. A. Kassavetis, and E. P. Geiduschek. 1992. A transcriptional enhancer whose function imposes

a requirement that proteins track along DNA. Science 256:1298–1303.

160. Herendeen, D. R., K. P. Williams, G. A. Kassavetis, and E. P. Geiduschek. 1990. An RNA polymerase-binding protein that is required for communication between an enhancer and a promoter. Science 248:573–578.
161. Herman, R. E., N. Haas, and D. P. Snustad. 1984. Identification of the bacteriophage T4 *unf* (= *alc*) gene product, a protein involved in the shutoff of host transcription. Genetics 108:305–317.
162. Herrmann, R. 1982. Nucleotide sequence of the bacteriophage T4 gene *57* and a deduced amino acid sequence. Nucleic Acids Res. 10:1105–1112.
163. Hershey, A. D., and M. Chase. 1952. Independent functions of viral protein and nucleic acid in growth of bacteriophage. J. Gen. Physiol. 36:39–56.
164. Hershey, A. D., and R. Rotman. 1949. Genetic recombination between host range and plaque-type mutants of bacteria in single bacterial cells. Genetics 34:44–71.
165. Hershey, A. D., and R. Rotman. 1948. Linkage among genes controlling inhibition of lysis in a bacterial virus. Proc. Natl. Acad. Sci. USA 34:89–96.
166. Hinton, D. M., R. March-Amegadzie, J. S. Gerber, and M. Sharma. 1996. Bacteriophage T4 middle transcription system: T4-modified RNA polymerase; AsiA, a sigma 70 binding protein; and transcriptional activator MotA. Methods Enzymol. 274:43–57.
167. Hinton, D. M., R. March-Amegadzie, J. S. Gerber, and M. Sharma. 1996. Characterization of pre-transcription complexes made at a bacteriophage T4 middle promoter: involvement of the T4 MotA activator and the T4 AsiA protein, a sigma 70 binding protein, in the formation of the open complex. J. Mol. Biol. 256:235–248.
168. Hinton, D. M., and S. Vuthoori. 2000. Efficient inhibition of *Escherichia coli* RNA polymerase by the bacteriophage T4 AsiA protein requires that AsiA binds first to free sigma70. J. Mol. Biol. 304:731–739.
169. Hobbs, L. J., and N. G. Nossal. 1996. Either bacteriophage T4 RNase H or *Escherichia coli* DNA polymerase I is essential for phage replication. J. Bacteriol. 178: 6772–6777.
170. Hsieh, P., C. S. Camerini-Otero, and R. D. Camerini-Otero. 1992. The synapsis event in the homologous pairing of DNAs: RecA recognizes and pairs less than one helical repeat of DNA. Proc. Natl. Acad. Sci. USA 89:6492–6496.
171. Huang, W. M., S. Z. Ao, S. Casjens, R. Orlandi, R. Zeikus, R. Weiss, D. Winge, and M. Fang. 1988. A persistent untranslated sequence within bacteriophage T4 DNA topoisomerase gene *60*. Science 239:1005–1012.
172. Hubermann, J. A. 1969. Visualization of replicating mammalian and T4 bacteriophage DNA. Cold Spring Harbor Symp. Quant. Biol. 33:509–524.
173. Huff, A. C., J. K. Leatherwood, and K. N. Kreuzer. 1989. Bacteriophage T4 DNA topoisomerase is the target of antitumor agent 4′-(9-acridinylamino)methanesulfon-m-anisidide (m-AMSA) in T4-infected *Escherichia coli*. Proc. Natl. Acad. Sci. USA 86:1307–1311.
174. Ishihama, A. 1992. Role of the RNA polymerase alpha subunit in transcription activation. Mol. Microbiol. 6:3283–3288.
175. Ishii, T., and M. Yanagida. 1977. The two dispensable structural proteins (soc and hoc) of the T4 phage capsid: their purification and properties, isolation and characterization of the defective mutants, and their binding with the defective heads in vitro. J. Mol. Biol. 109: 487–514.
176. Iwasaki, K., B. L. Trus, P. T. Wingfield, N. Cheng, G. Campusano, V. B. Rao, and A. C. Steven. 2000. Molecular architecture of bacteriophage T4 capsid: vertex structure and bimodal binding of the stabilizing accessory protein Virology 271:321–333.
177. Jardine, P. J., and D. H. Coombs. 1998. Capsid expansion follows the initiation of DNA packaging in bacteriophage T4. J. Mol. Biol. 284:661–672.
178. Jardine, P. J., M. C. McCormick, C. Lutze-Wallace, and D. H. Coombs. 1998. The bacteriophage T4 DNA packaging apparatus targets the unexpanded prohead. J. Mol. Biol. 284:647–659.
179. Jones, C. E., T. C. Mueser, K. C. Dudas, K. N. Kreuzer, and N. G. Nossal. 2001. Bacteriophage T4 gene *41* helicase and gene *59* helicase-loading protein: a versatile couple with roles in replication and recombination. Proc. Natl. Acad. Sci. USA 98:8312–8318.
180. Jones, C. E., T. C. Mueser, and N. G. Nossal. 2000. Interaction of the bacteriophage T4 gene *59* helicase loading protein and gene *41* helicase with each other and with fork, flap, and cruciform DNA. J. Biol. Chem. 275: 27145–27154.
181. Kadyrov, F. A., and J. W. Drake. 2001. Conditional coupling of leading-strand and lagging-strand DNA synthesis at bacteriophage T4 replication forks. J. Biol. Chem. 276: 29559–29566.
182. Kadyrov, F. A., and V. M. Kriukov. 1996. [Phage T4 segE gene “mobility”: high frequency of *segE* gene transfer from the plasmid into the phage genome depends on the intactness of this gene]. Dokl. Akad. Nauk. 346: 700–704.
183. Kadyrov, F. A., M. G. Shlyapnikov, and V. M. Kryukov. 1997. A phage T4 site-specific endonuclease, SegE, is responsible for a non-reciprocal genetic exchange between T-even-related phages. FEBS Lett. 415:75–80.
184. Kai, T., H. E. Selick, and T. Yonesaki. 1996. Destabilization of bacteriophage T4 mRNAs by a mutation of gene *61.5*. Genetics 144:7–14.
185. Kai, T., H. Ueno, and T. Yonesaki. 1998. Involvement of other bacteriophage T4 genes in the blockade of protein synthesis and mRNA destabilization by a mutation of gene *61.5*. Virology 248:148–155.
186. Kai, T., and T. Yonesaki. 2002. Multiple mechanisms for degradation of bacteriophage T4 soc mRNA. Genetics 160:5–12.
187. Kanamaru, S., P. G. Leiman, V. A. Kostyuchenko, P. R. Chipman, V. V. Mesyanzhinov, F. Arisaka, and M. G. Rossmann. 2002. Structure of the cell-puncturing device of bacteriophage T4. Nature 415:553–557.
188. Kang, C., R. Chan, I. Berger, C. Lockshin, L. Green, L. Gold, and A. Rich. 1995. Crystal structure of the T4 regA translational regulator protein at 1.9 Å resolution. Science 268:1170–1173.
189. Kao, S. H., and W. H. McClain. 1980. Baseplate protein of bacteriophage T4 with both structural and lytic functions. J. Virol. 34:95–103.

190. Kao, S. H., and W. H. McClain. 1980. Roles of bacteriophage T4 gene *5* and gene *s* products in cell lysis. J. Virol. 34: 104–107.
191. Karam, J. D., and B. Barker. 1971. Properties of bacteriophage T4 mutants defective in gene *30* (deoxyribonucleic acid ligase) and the *rII* gene. J. Virol. 7:260–266.
192. Karam, J. D., J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, C. Carlson, and E. S. Miller. (eds.) 1994. Molecular Biology of Bacteriophage T4. American Society for Microbiology, Washington, D.C.
193. Karam, J. D., and W. H. Konigsberg. 2000. DNA polymerase of the T4-related bacteriophages. Prog. Nucleic Acid Res. Mol. Biol. 64:65–96.
194. Kashlev, M., E. Nudler, A. Goldfarb, T. White, and E. Kutter. 1993. Bacteriophage T4 Alc protein: a transcription termination factor sensing local modification of DNA. Cell 75:147–154.
195. Kellenberger, E. 1980. Control mechanisms in the morphogeneses of bacteriophage heads. Biosystems 12: 201–223.
196. Kellenberger, E., and E. Boy de la Tour. 1964. On the fine structure of normal and "polymerized" tail sheath of phage T4. J. Ultrastruct. Res. 11:545–563.
197. Kells, S. S., and R. Haselkorn. 1974. Bacteriophage T4 short tail fibers are the product of gene *12*. J. Mol. Biol. 83:473–485.
198. Kemper, B. 1998. Branched DNA resolving enzymes (X-solvases), pp. 179–204. *In* J. A. Nickoloff and M. Hoekstra (eds.) DNA Damage and Repair, vol. 1: DNA Repair in Prokaryotes and Lower Eukaryotes. Humana Press, Totowa, N.J.
199. Kemper, B., and D. T. Brown. 1976. Function of gene *49* of bacteriophage T4. II. Analysis of intracellular development and the structure of very fast-sedimenting DNA. J. Virol. 18:1000–1015.
200. Kemper, B., F. Jensch, M. von Depka-Prondzynski, H. J. Fritz, U. Borgmeyer, and K. Mizuuchi. 1984. Resolution of Holliday structures by endonuclease VII as observed in interactions with cruciform DNA. Cold Spring Harb. Symp. Quant. Biol. 49:815–825.
201. Kim, J. S., and N. Davidson. 1974. Electron microscope heteroduplex study of sequence relations of T2, T4, and T6 bacteriophage DNAs. Virology 57:93–111.
202. King, G. J., and W. M. Huang. 1982. Identification of the origins of T4 DNA replication. Proc. Natl. Acad. Sci. USA 79:7248–7252.
203. Kirschner, M., and J. Gerhart. 1998. Evolvability. Proc. Natl. Acad. Sci. USA 95:8420–8427.
204. Kleff, S., and B. Kemper. 1988. Initiation of heteroduplex-loop repair by T4-encoded endonuclease VII in vitro. EMBO J. 7:1527–1535.
205. Klein, H. L., and K. N. Kreuzer. 2002. Replication, recombination, and repair: going for the gold. Mol. Cell 9:471–480.
206. Koch, T., A. Raudonikiene, K. Wilkens, and W. Ruger. 1995. Overexpression, purification, and characterization of the ADP-ribosyltransferase (gpAlt) of bacteriophage T4: ADP-ribosylation of *E. coli* RNA polymerase modulates T4 "early" transcription. Gene Expr. 4:253–264.
207. Kocsis, E., H. L. Greenstone, E. G. Locke, M. Kessel, and A. C. Steven. 1997. Multiple conformational states of the bacteriophage T4 capsid surface lattice induced when expansion occurs without prior cleavage. J. Struct. Biol. 118:73–82.
208. Kogoma, T. 1997. Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. Microbiol. Mol. Biol. Rev. 61: 212–238.
209. Kolesky, S., M. Ouhammouch, E. N. Brody, and E. P. Geiduschek. 1999. Sigma competition: the contest between bacteriophage T4 middle and late transcription. J. Mol. Biol. 291:267–281.
210. Kornberg, A., and T. A. Baker. 1992. DNA replication, 2nd edn. W. H. Freeman, New York.
211. Kozinski, A. W. 1983. Origins of T4 DNA replication, pp. 111–119. *In* C. K. Mathews, E. Kutter, G. Mosig, and P. B. Berget (eds.) Bacteriophage T4. American Society for Microbiology, Washington, D.C.
212. Kreuzer, K. N. 1994. A bacteriophage model system for studying topoisomerase inhibitors. Adv. Pharmacol. 29B:171–186.
213. Kreuzer, K. N. 1998. Bacteriophage T4, a model system for understanding the mechanism of type II topoisomerase inhibitors. Biochim. Biophys. Acta 1400:339–347.
214. Kreuzer, K. N. 2000. Recombination-dependent DNA replication in phage T4. Trends Biochem. Sci. 25:165–173.
215. Kreuzer, K. N., and B. M. Alberts. 1986. Characterization of a defective phage system for the analysis of bacteriophage T4 DNA replication origins. J. Mol. Biol. 188: 185–198.
216. Kreuzer, K. N., and B. M. Alberts. 1985. A defective phage system reveals bacteriophage T4 replication origins that coincide with recombination hot spots. Proc. Natl. Acad. Sci. USA 82:3345–3349.
217. Kreuzer, K. N., and J. W. Drake. 1994. Repair of lethal DNA damage, pp. 89–97. *In* J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, C. Carlson, and E. S. Miller (eds.) Molecular Biology of Bacteriophage T4. American Society for Microbiology, Washington, D.C.
218. Kreuzer, K. N., and S. W. Morrical. 1994. Initiation of DNA replication, pp. 28–42. *In* J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, C. Carlson, and E. S. Miller (eds.) Molecular Biology of Bacteriophage T4. American Society for Microbiology, Washington, D.C.
219. Kreuzer, K. N., M. Saunders, L. J. Weislo, and H. W. Kreuzer. 1995. Recombination-dependent DNA replication stimulated by double-strand breaks in bacteriophage T4. J. Bacteriol. 177:6844–6853.
220. Krisch, H. M., and B. Allet. 1982. Nucleotide sequences involved in bacteriophage T4 gene 32 translational self-regulation. Proc. Natl. Acad. Sci. USA 79:4937–4941.
221. Krisch, H. M., D. B. Shah, and H. Berger. 1971. Replication and recombination in ligase-deficient rII bacteriophage T4D. J. Virol. 7:491–498.
222. Krisch, H. M., G. Van Houwe, D. Belin, W. Gibbs, and R. H. Epstein. 1977. Regulation of the expression of bacteriophage T4 genes *32* and *43*. Virology 78:87–98.

223. Kuebler, D., and V. B. Rao. 1998. Functional analysis of the DNA-packaging/terminase protein gp17 from bacteriophage T4. J. Mol. Biol. 281:803–814.
224. Kunisawa, T., S. Kanaya, and E. Kutter. 1998. Comparison of synonymous codon distribution patterns of bacteriophage and host genomes. DNA Res. 5:319–326.
225. Kutter, E. 1996. Analysis of bacteriophage T4 based on the completed sequence data, pp. 13–28. *In* J. Collado-Vides, B. Magasanik, and T. Smith (eds.) Integrative Approaches to Molecular Biology. MIT Press, Cambridge, Mass.
226. Kutter, E., K. Gachechiladze, A. Poglazov, E. Marusich, M. Shneider, P. Aronsson, A. Napuli, D. Porter, and V. Mesyanzhinov. 1995. Evolution of T4-related phages. Virus Genes 11:285–297.
227. Kutter, E., B. Guttman, and K. Carlson. 1994. The transition from host to phage metabolism after T4 infection, pp. 343–346. *In* J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, C. Carlson, and E. S. Miller (eds.) Molecular Biology of Bacteriophage T4. American Society for Microbiology, Washington, D.C.
228. Kutter, E., T. White, M. Kashlev, M. Uzan, J. McKinney, and B. Guttman. 1994. Effects on host genome structure and expression, pp. 357–368. *In* J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, C. Carlson, and E. S. Miller (eds.) Molecular Biology of Bacteriophage T4. American Society for Microbiology, Washington, D.C.
229. Kutter, E. M., and J. S. Wiberg. 1968. Degradation of cytosine-containing bacterial and bacteriophage DNA after infection of *Escherichia coli* B with bacteriophage T4D wild type and with mutants defective in genes *46*, *47* and *56*. J. Mol. Biol. 38:395–411.
230. Lambert, L. J., V. Schirf, B. Demeler, M. Cadene, and M. H. Werner. 2001. Flipping a genetic switch by subunit exchange. EMBO J. 20:7149–7159.
231. Lane, T., and F. Eiserling. 1990. Genetic control of capsid length in bacteriophage T4. VII. A model of length regulation based on DNA size. J. Struct. Biol. 104:9–23.
232. Lane, T., P. Serwer, S. J. Hayes, and F. Eiserling. 1990. Quantized viral DNA packaging revealed by rotating gel electrophoresis. Virology 174:472–478.
233. Lawrence, J. G., G. F. Hatfull, and R. W. Hendrix. 2002. Imbroglios of viral taxonomy: genetic exchange and failings of phenetic approaches. J. Bacteriol. 184: 4891–4905.
234. Leffers, G., and V. B. Rao. 2000. Biochemical characterization of an ATPase activity associated with the large packaging subunit gp17 from bacteriophage T4. J. Biol. Chem. 275:37127–37136.
235. Leibo, S. P., E. Kellenberger, C. Kellenberger-van der Kamp, T. G. Frey, and C. M. Steinberg. 1979. Gene *24*-controlled osmotic shock resistance in bacteriophage T4: probable multiple gene functions. J. Virol. 30:327–338.
236. Leiman, P. G., V. A. Kostyuchenko, M. M. Shneider, L. P. Kurochkina, V. V. Mesyanzhinov, and M. G. Rossmann. 2000. Structure of bacteriophage T4 gene product 11, the interface between the baseplate and short tail fibers. J. Mol. Biol. 301:975–985.
237. Lepault, J., and K. Leonard. 1985. Three-dimensional structure of unstained, frozen-hydrated extended tails of bacteriophage T4. J. Mol. Biol. 182:431–441.
238. Liang, Y. M., R. X. Wei, T. Hsu, C. Alford, M. Dawson, and J. Karam. 1988. Autogenous regulation of the *regA* gene of bacteriophage T4: derepression of translation. Genetics 119:743–749.
239. Liebig, H. D., and W. Ruger. 1989. Bacteriophage T4 early promoter regions. Consensus sequences of promoters and ribosome-binding sites. J. Mol. Biol. 208:517–536.
240. Lin, G. W. 1988. PhD thesis, Vanderbilt University.
241. Lin, H., and L. W. Black. 1998. DNA requirements in vivo for phage T4 packaging. Virology 242:118–127.
242. Lin, H., V. B. Rao, and L. W. Black. 1999. Analysis of capsid portal protein and terminase functional domains: interaction sites required for DNA packaging in bacteriophage T4. J. Mol. Biol. 289:249–260.
243. Lin, H., M. N. Simon, and L. W. Black. 1997. Purification and characterization of the small subunit of phage T4 terminase, gp16, required for DNA packaging. J. Biol. Chem. 272:3495–3501.
244. Lipinska, B., A. S. Rao, B. M. Bolten, R. Balakrishnan, and E. B. Goldberg. 1989. Cloning and identification of bacteriophage T4 gene 2 product gp2 and action of gp2 on infecting DNA in vivo. J. Bacteriol. 171:488–497.
245. Lloyd, R. S. 1999. The initiation of DNA base excision repair of dipyrimidine photoproducts. Prog. Nucleic Acid Res. Mol. Biol. 62:155–175.
246. Luder, A., and G. Mosig. 1982. Two alternative mechanisms for initiation of DNA replication forks in bacteriophage T4: priming by RNA polymerase and by recombination. Proc. Natl. Acad. Sci. USA 79:1101–1105.
247. Luria, S. E., and M. L. Human. 1952. A nonhereditary, host-induced variation of bacterial viruses. J. Bacteriol. 64:557–569.
248. Macdonald, P. M., R. M. Seaby, W. Brown, and G. Mosig (eds.). 1983. Initiator DNA from a primary origin and induction of a secondary origin of bacteriophage T4 DNA replication, pp. 111–116. In D. Schlessinger (ed.). Microbiology 1983. American Society for Microbiology Washington, D. C.
249. Makhov, A. M., B. L. Trus, J. F. Conway, M. N. Simon, T. G. Zurabishvili, V. V. Mesyanzhinov, and A. C. Steven. 1993. The short tail-fiber of bacteriophage T4: molecular structure and a mechanism for its conformational transition. Virology 194:117–127.
250. Manne, V., V. B. Rao, and L. W. Black. 1982. A bacteriophage T4 DNA packaging related DNA-dependent ATPase-endonuclease. J. Biol. Chem. 257:13223–13232.
251. March-Amegadzie, R., and D. M. Hinton. 1995. The bacteriophage T4 middle promoter PuvsX: analysis of regions important for binding of the T4 transcriptional activator MotA and for activation of transcription. Mol. Microbiol. 15:649–660.
252. Marsh, R. C., A. M. Breschkin, and G. Mosig. 1971. Origin and direction of bacteriophage T4 DNA replication. II. A gradient of marker frequencies in partially replicated T4 DNA as assayed by transformation. J. Mol. Biol. 60: 213–233.

253. Mathews, C. K., E. Kutter, G. Mosig, and P. B. Berget (eds.). 1983. Bacteriophage T4. American Society for Microbiology, Washington, D.C.
254. Matthews, B. W., and S. J. Remington. 1974. The three dimensional structure of the lysozyme from bacteriophage T4. Proc. Natl. Acad. Sci. USA 71:4178–4182.
255. Mattson, T., J. Richardson, and D. Goodin. 1974. Mutant of bacteriophage T4D affecting expression of many early genes. Nature 250:48–50.
256. McNicol, L. A., and L. E. Simon. 1977. A mutation which bypasses the requirement for p24 in bacteriophage T4 capsid morphogenesis. J. Mol. Biol. 116:261–283.
257. McPheeters, D. S., G. D. Stormo, and L. Gold. 1988. Autogenous regulatory site on the bacteriophage T4 gene 32 messenger RNA. J. Mol. Biol. 201:517–535.
258. Meselson, M., and J. J. Weigle. 1961. Chromosome brekage accompanying genetic recombination in bacteriophage. Proc. Natl. Acad. Sci. USA 47:857–868.
259. Michel, C. J., B. Jacq, D. G. Arques, and T. A. Bickle. 1986. A remarkable amino acid sequence homology between a phage T4 tail fibre protein and ORF314 of phage lambda located in the tail operon. Gene 44:147–150.
260. Miller, E. S., J. F. Heidelberg, J. A. Eisen, W. C. Nelson, A. S. Durkin, A. Ciecko, T. V. Feldblyum, O. White, I. T. Paulsen, W. C. Nierman, J. Lee, B. Szczypinski, and C. M. Fraser. 2003. Complete genome sequence of the broad-host-range vibriophage KVP40: comparative genomics of a T4-related bacteriophage. J. Bacteriol. 185: 5220–5233.
261. Miller, E. S., J. D. Karam, and E. Spicer. 1994. Control of translational initiation: mRNA structure and protein repressors, pp. 193–208. *In* J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, C. Carlson, and E. S. Miller (eds.) Molecular Biology of Bacteriophage T4. American Society for Microbiology, Washington, D.C.
262. Miller, E. S., E. Kutter, G. Mosig, F. Arisaka, T. Kunisawa, and W. Ruger. 2003. Bacteriophage T4 genome. Microbiol. Mol. Biol. Rev. 67:86–156.
263. Miller, E. S., R. B. Winter, K. M. Campbell, S. D. Power, and L. Gold. 1985. Bacteriophage T4 regA protein. Purification of a translational repressor. J. Biol. Chem. 260:13053–13059.
264. Minagawa, T., H. Fujisawa, T. Yonesaki, and Y. Ryo. 1988. Function of cloned T4 recombination genes, *uvsX* and *uvsY*, in cells of *Escherichia coli*. Mol. Gen. Genet. 211: 350–356.
265. Minakhin, L., J. A. Camarero, M. Holford, C. Parker, T. W. Muir, and K. Severinov. 2001. Mapping the molecular interface between the sigma(70) subunit of *E. coli* RNA polymerase and T4 AsiA. J. Mol. Biol. 306:631–642.
266. Monod, C., F. Repoila, M. Kutateladze, F. Tetart, and H. M. Krisch. 1997. The genome of the pseudo T-even bacteriophages, a diverse group that resembles T4. J. Mol. Biol. 267:237–249.
267. Moody, M. F. 1999. Geometry of phage head construction. J. Mol. Biol. 293:401–433.
268. Moody, M. F., and L. Makowski. 1981. X-ray diffraction study of tail-tubes from bacteriophage T2L. J. Mol. Biol. 150:217–244.
269. Morais, M. C., Y. Tao, N. H. Olson, S. Grimes, P. J. Jardine, D. L. Anderson, T. S. Baker, and M. G. Rossmann. 2001. Cryoelectron-microscopy image reconstruction of symmetry mismatches in bacteriophage phi29. J. Struct. Biol. 135:38–46.
270. Morrical, S., K. Hempstead, M. Morrical, K. M. Chou, R. Ando, and O. Grigorieva. 1994. Mechanisms of assembly of the enzyme–ssDNA complexes required for recombination-dependent DNA synthesis and repair in bacteriophage T4. Ann. N. Y. Acad. Sci. 726:349–350.
271. Mosig, G. 1985. Bacteriophage T4 gene 32 participates in excision repair as well as recombinational repair of UV damages. Genetics 110:159–171.
272. Mosig, G. 1963. Coordinate variation in density and recombination potential in T4 phage particles produced at different times after infection. Genetics 48:1195–1200.
273. Mosig, G. 1966. Distances separating genetic markers in T4 DNA. Proc. Natl. Acad. Sci. USA 56:1177–1183.
274. Mosig, G. 1987. The essential role of recombination in phage T4 growth. Annu. Rev. Genet. 21:347–371.
275. Mosig, G. (ed.). 1978. Evidence for a Complex that Coordinates Different Steps in General Genetic Recombination. Springer, Berlin Germany.
276. Mosig, G. 1963. Genetic recombination in bacteriophage T4 during replication of DNA fragments. Cold Spring Harbor Symp. Quant. Biol. 28:35–42.
277. Mosig, G. 1994. Homologous recombination, pp. 54–82. *In* J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, C. Carlson, and E. S. Miller (eds.) Molecular Biology of Bacteriophage T4. American Society for Microbiology, Washington, D.C.
278. Mosig, G. 1968. A map of distances along the DNA molecule of phage T4. Genetics 59:137–151.
279. Mosig, G. 1998. Recombination and recombination-dependent DNA replication in bacteriophage T4. Annu. Rev. Genet. 32:379–413.
280. Mosig, G. 1983. Relationship of T4 DNA replication and recombination, pp. 120–130. *In* C. K. Mathews, E. Kutter, G. Mosig, and P. B. Berget (eds.) Bacteriophage T4. American Society for Microbiology, Washington, D.C.
281. Mosig, G. 1994. Synthesis and maturation of T4-encoded tRNAs, pp. 182–185. *In* J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, C. Carlson, and E. S. Miller (eds.) Molecular Biology of Bacteriophage T4. American Society for Microbiology, Washington, D.C.
282. Mosig, G., S. Benedict, D. Ghosal, A. Luder, R. Dannenberg, and S. Bock. 1980. Genetic Analysis of DNA Replication in Bacteriophage T4. Mechanistic Studies of DNA Replication and Genetic Recombination. Academic Press, Nashville, Tenn.
283. Mosig, G., and S. Bock. 1976. Gene *32* protein of bacteriophage T4 moderates the activities of the T4 gene *46/47*-controlled nuclease and of the *Escherichia coli* RecBC nuclease in vivo. J. Virol. 17:756–761.
284. Mosig, G., and A. M. Breschkin. 1975. Genetic evidence for an additional function of phage T4 gene *32* protein: interaction with ligase. Proc. Natl. Acad. Sci. USA 72: 1226–1230.

285. Mosig, G., and R. Calendar. 2001. Horizontal gene transfer in bacteriophages, chapter 13. *In* M. Syvanen (ed.) Horizontal Gene Transfer. Academic Press, London.
286. Mosig, G., and N. Colowick. 1995. DNA replication of bacteriophage T4 in vivo, pp. 587–604. *In* J. Campbell (eds.) Methods in Enzymology, vol. 262. Academic Press, San Diego.
287. Mosig, G., N. Colowick, M. E. Gruidl, A. Chang, and A. J. Harvey. 1995. Multiple initiation mechanisms adapt phage T4 DNA replication to physiological changes during T4's development. FEMS Microbiol. Rev. 17:83–98.
288. Mosig, G., N. E. Colowick, and B. C. Pietz. 1998. Several new bacteriophage T4 genes, mapped by sequencing deletion endpoints between genes *56* (dCTPase) and *dda* (a DNA-dependent ATPase-helicase) modulate transcription. Gene 223:143–155.
289. Mosig, G., R. Ehring, W. Schliewen, and S. Bock. 1971. The patterns of recombination and segregation in terminal regions of T4DNA molecules. Mol. Gen. Genet. 113:51–91.
290. Mosig, G., J. Gewin, A. Luder, N. Colowick, and D. Vo. 2001. Two recombination-dependent DNA replication pathways of bacteriophage T4, and their roles in mutagenesis and horizontal gene transfer. Proc. Natl. Acad. Sci. USA 98: 8306–8311.
291. Mosig, G., and D. H. Hall. 1994. Gene Expression: a paradigm of integrated circuits, pp. 127–131. *In* J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L.W. Black, E. K. Spicer, E. Kutter, C. Carlson, and E. S. Miller (eds.) Molecular Biology of Bacteriophage T4. American Society for Microbiology, Washington, D.C.
292. Mosig, G., G. W. Lin, J. Franklin, and W. H. Fan. 1989. Functional relationships and structural determinants of two bacteriophage T4 lysozymes: a soluble (gene *e*) and a baseplate-associated (gene *5*) protein. New Biol. 1:171–179.
293. Mosig, G., A. Luder, A. Ernst, and N. Canan. 1991. Bypass of a primase requirement for bacteriophage T4 DNA replication in vivo by a recombination enzyme, endonuclease VII. New Biol. 3:1195–1205.
294. Mosig, G., P. Macdonald, G. Lin, M. Levin, and R. Seaby. 1983. Gene expression and initiation of DNA replication of bacteriophage T4 in phage and host topoisomerase mutants, pp. 73–186. *In* N. R. Cozzarelli (ed.) Mechanisms of DNA replication and recombination. A. R. Liss, New York.
295. Mosig, G., and D. Powell. 1985. Heteroduplex loops are packaged in gene *49* (endonuclease VII) mutants of bacteriophage T4. Abstr. Annu. Meet. Am. Soc. Microbiol. 85:209.
296. Mosig, G., M. Shaw, and G. M. Garcia. 1984. On the role of DNA replication, endonuclease VII, and rII proteins in processing of recombinational intermediates in phage T4. Cold Spring Harb. Symp. Quant. Biol. 49:371–382.
297. Mosig, G., S. Yu, H. Myung, E. Haggard-Ljungquist, L. Davenport, K. Carlson, and R. Calendar. 1997. A novel mechanism of virus–virus interactions: bacteriophage P2 Tin protein inhibits phage T4 DNA synthesis by poisoning the T4 single-stranded DNA binding protein, gp32. Virology 230:72–81.
298. Mudd, E. A., H. M. Krisch, and C. F. Higgins. 1990. RNase E, an endoribonuclease, has a general role in the chemical decay of *Escherichia coli* mRNA: evidence that rne and ams are the same genetic locus. Mol. Microbiol. 4: 2127–2135.
299. Mueller, J. E., J. Clyman, Y. J. Huang, M. M. Parker, and M. Belfort. 1996. Intron mobility in phage T4 occurs in the context of recombination-dependent DNA replication by way of multiple pathways. Genes Dev. 10:351–364.
300. Mueller, J. E., D. Smith, and M. Belfort. 1996. Exon coconversion biases accompanying intron homing: battle of the nucleases. Genes Dev. 10:2158–2166.
301. Mueser, T. C., C. E. Jones, N. G. Nossal, and C. C. Hyde. 2000. Bacteriophage T4 gene *59* helicase assembly protein binds replication fork DNA. The 1.45 Å resolution crystal structure reveals a novel alpha-helical two-domain fold. J. Mol. Biol. 296:597–612.
302. Mueser, T. C., N. G. Nossal, and C. C. Hyde. 1996. Structure of bacteriophage T4 RNase H, a 5′ to 3′ RNA–DNA and DNA–DNA exonuclease with sequence similarity to the RAD2 family of eukaryotic proteins. Cell 85:1101–1112.
303. Muller-Salamin, L., L. Onorato, and M. K. Showe. 1977. Localization of minor protein components of the head of bacteriophage T4. J. Virol. 24:121–134.
304. Nakagawa, H., F. Arisaka, and S. Ishii. 1985. Isolation and characterization of the bacteriophage T4 tail-associated lysozyme. J. Virol. 54:460–466.
305. Nassif, N., J. Penney, S. Pal, W. R. Engels, and G. B. Gloor. 1994. Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. Mol. Cell Biol. 14:1613–1625.
306. Neece, S. H., K. Carles-Kinch, D. J. Tomso, and K. N. Kreuzer. 1996. Role of recombinational repair in sensitivity to an antitumour agent that inhibits bacteriophage T4 type II DNA topoisomerase. Mol. Microbiol. 20:1145–1154.
307. Nossal, N. G. 1994. The bacteriophage T4 replication fork, pp. 43–53. *In* J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, C. Carlson, and E. S. Miller (eds.) Molecular Biology of Bacteriophage T4. American Society for Microbiology, Washington, D.C.
308. Nossal, N. G. 1998. A new look at old mutants of T4 DNA polymerase. Genetics 148:1535–1538.
309. Nossal, N. G., K. C. Dudas, and K. N. Kreuzer. 2001. Bacteriophage T4 proteins replicate plasmids with a preformed R loop at the T4 ori(uvsY) replication origin in vitro. Mol. Cell 7:31–41.
310. Nossal, N. G., D. M. Hinton, L. J. Hobbs, and P. Spacciapoli. 1995. Purification of bacteriophage T4 DNA replication proteins. Methods Enzymol. 262:560–584.
311. Olson, N. H., M. Gingery, F. A. Eiserling, and T. S. Baker. 2001. The structure of isometric capsids of bacteriophage T4. Virology 279:385–391.
312. O'Malley, S. M., A. K. Sattar, K. R. Williams, and E. K. Spicer. 1995. Mutagenesis of the COOH-terminal region of bacteriophage T4 regA protein. J. Biol. Chem. 270:5107–5114.
313. Ouhammouch, M., K. Adelman, S. R. Harvey, G. Orsini, and E. N. Brody. 1995. Bacteriophage T4 MotA and AsiA proteins suffice to direct *Escherichia coli* RNA polymerase

to initiate transcription at T4 middle promoters. Proc. Natl. Acad. Sci. USA 92:1451–1455.

314. Ouhammouch, M., G. Orsini, and E. N. Brody. 1994. The *asiA* gene product of bacteriophage T4 is required for middle mode RNA synthesis. J. Bacteriol. 176: 3956–3965.

315. Paddison, P., S. T. Abedon, H. K. Dressman, K. Gailbreath, J. Tracy, E. Mosser, J. Neitzel, B. Guttman, and E. Kutter. 1998. The roles of the bacteriophage T4 *r* genes in lysis inhibition and fine-structure genetics: a new perspective. Genetics 148:1539–1550.

316. Pahari, S., and D. Chatterji. 1997. Interaction of bacteriophage T4 AsiA protein with *Escherichia coli* sigma70 and its variant. FEBS Lett. 411:60–62.

317. Panne, D., S. A. Muller, S. Wirtz, A. Engel, and T. A. Bickle. 2001. The McrBC restriction endonuclease assembles into a ring structure in the presence of G nucleotides. EMBO J. 20:3210–3217.

318. Panne, D., E. A. Raleigh, and T. A. Bickle. 1998. McrBs, a modulator peptide for McrBC activity. EMBO J. 17: 5477–5483.

319. Paques, F., and J. E. Haber. 1999. Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. Microbiol. Mol. Biol. Rev. 63: 349–404.

320. Parker, M. M., M. Belisle, and M. Belfort. 1999. Intron homing with limited exon homology. Illegitimate double-strand-break repair in intron acquisition by phage T4. Genetics 153:1513–1523.

321. Parma, D. H., M. Snyder, S. Sobolevski, M. Nawroz, E. Brody, and L. Gold. 1992. The Rex system of bacteriophage lambda: tolerance and altruistic cell death. Genes Dev. 6:497–510.

322. Paulson, J. R., and U. K. Laemmli. 1977. Morphogenetic core of the bacteriophage T4 head. Structure of the core in polyheads. J. Mol. Biol. 111:459–485.

323. Pavlov, A. R., and J. D. Karam. 1994. Binding specificity of T4 DNA polymerase to RNA. J. Biol. Chem. 269: 12968–12972.

324. Pavlov, A. R., and J. D. Karam. 2000. Nucleotide-sequence-specific and non-specific interactions of T4 DNA polymerase with its own mRNA. Nucleic Acids Res. 28: 4657–4664.

325. Pene, C., and M. Uzan. 2000. The bacteriophage T4 anti-sigma factor AsiA is not necessary for the inhibition of early promoters in vivo. Mol. Microbiol. 35: 1180–1191.

326. Phillips, C. A., J. Gordon, and E. K. Spicer. 1996. Bacteriophage T4 regA protein binds RNA as a monomer, overcoming dimer interactions. Nucleic Acids Res. 24: 4319–4326.

327. Qiu, H., K. Kaluarachchi, Z. Du, D. W. Hoffman, and D. P. Giedroc. 1996. Thermodynamics of folding of the RNA pseudoknot of the T4 gene 32 autoregulatory messenger RNA. Biochemistry 35:4176–4186.

328. Raaijmakers, H., O. Vix, I. Toro, S. Golz, B. Kemper, and D. Suck. 1999. X-ray structure of T4 endonuclease VII: a DNA junction resolvase with a novel fold and unusual domain-swapped dimer architecture. EMBO J. 18: 1447–1458.

329. Ramanculov, E., and R. Young. 2001. An ancient player unmasked: T4 rI encodes a t-specific antiholin. Mol. Microbiol. 41:575–583.

330. Ramanculov, E., and R. Young. 2001. Functional analysis of the phage T4 holin in a lambda context. Mol. Genet. Genomics 265:345–353.

331. Rao, V. B., and L. W. Black. 1988. Cloning, overexpression and purification of the terminase proteins gp16 and gp17 of bacteriophage T4. Construction of a defined in-vitro DNA packaging system using purified terminase proteins. J. Mol. Biol. 200:475–488.

332. Rao, V. B., and M. S. Mitchell. 2001. The N-terminal ATPase site in the large terminase protein gp17 is critically required for DNA packaging in bacteriophage T4. J. Mol. Biol. 314:401–411.

333. Rao, V. B., V. Thaker, and L. W. Black. 1992. A phage T4 in vitro packaging system for cloning long DNA molecules. Gene 113:25–33.

334. Reha-Krantz, L. J. 1994. Genetic dissection of T4 DNA polymerase structure–function relationships, pp. 307–312. *In* J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, C. Carlson, and E. S. Miller (eds.) Molecular Biology of Bacteriophage T4. American Society for Microbiology, Washington, D.C.

335. Reha-Krantz, L. J. 1995. Learning about DNA polymerase function by studying antimutator DNA polymerases. Trends Biochem. Sci. 20:136–140.

336. Reha-Krantz, L. J. 1995. Use of genetic analyses to probe structure, function, and dynamics of bacteriophage T4 DNA polymerase. Methods Enzymol. 262:323–331.

337. Reha-Krantz, L. J., L. A. Marquez, E. Elisseeva, R. P. Baker, L. B. Bloom, H. B. Dunford, and M. F. Goodman. 1998. The proofreading pathway of bacteriophage T4 DNA polymerase. J. Biol. Chem. 273:22969–22976.

338. Revel, H. R. 1983. DNA modification: glucosylation, pp. 156–165. *In* C. K. Mathews, E. Kutter, G. Mosig, and P. B. Berget (eds.) Bacteriophage T4. American Society for Microbiology, Washington, D.C.

339. Riede, I., K. Drexler, M. L. Eschbach, and U. Henning. 1986. DNA sequence of the tail fiber genes *37*, encoding the receptor recognizing part of the fiber, of bacteriophages T2 and K3. J. Mol. Biol. 191:255–266.

340. Rodriguez-Prieto, A. 1976. PhD thesis, Vanderbilt University.

341. Rosario, M. O., and J. W. Drake. 1990. Frameshift and double-amber mutations in the bacteriophage T4 *uvsX* gene: analysis of mutant UvsX proteins from infected cells. Mol. Gen. Genet. 222:112–119.

342. Ross, P. D., L. W. Black, M. E. Bisher, and A. C. Steven. 1985. Assembly-dependent conformational changes in a viral capsid protein. Calorimetric comparison of successive conformational states of the gp23 surface lattice of bacteriophage T4. J. Mol. Biol. 183:353–364.

343. Ross, W., K. K. Gosink, J. Salomon, K. Igarashi, C. Zou, A. Ishihama, K. Severinov, and R. L. Gourse. 1993. A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. Science 262:1407–1413.

344. Rossmann, M. G. 2000. Fitting atomic models into electron-microscopy maps. Acta Crystallogr. D Biol. Crystallogr. 56:1341–1349.

345. Rossmann, M. G., V. V. Mesyanzhinov, F. Arisaka, and P. G. Leiman. 2004. The bacteriophage T4 DNA injection machine. Curr. Opin. Struct. Biol. 14:171–180.

346. Ruckman, J., D. Parma, C. Tuerk, D. H. Hall, and L. Gold. 1989. Identification of a T4 gene required for bacteriophage mRNA processing. New Biol. 1:54–65.

347. Ruckman, J., S. Ringquist, E. Brody, and L. Gold. 1994. The bacteriophage T4 regB ribonuclease. Stimulation of the purified enzyme by ribosomal protein S1. J. Biol. Chem. 269:26655–26662.

348. Salinas, F., and T. Kodadek. 1995. Phage T4 homologous strand exchange: a DNA helicase, not the strand transferase, drives polar branch migration. Cell 82: 111–119.

349. Sanders, G. M., G. A. Kassavetis, and E. P. Geiduschek. 1997. Dual targets of a transcriptional activator that tracks on DNA. EMBO J. 16:3124–3132.

350. Sanders, G. M., G. A. Kassavetis, and E. P. Geiduschek. 1995. Rules governing the efficiency and polarity of loading a tracking clamp protein onto DNA: determinants of enhancement in bacteriophage T4 late transcription. EMBO J. 14:3966–3976.

351. Sanson, B., and M. Uzan. 1995. Post-transcriptional controls in bacteriophage T4: roles of the sequence-specific endoribonuclease RegB. FEMS Microbiol. Rev. 17:141–150.

352. Schmidt, F. J., and D. Apirion. 1983. T4 transfer RNAs: paradigmatic system for the study of RNA processing, pp. 208–217. *In* C. K. Mathews, E. Kutter, G. Mosig, and P. B. Berget (eds.) Bacteriophage T4. American Society for Microbiology, Washington, D.C.

353. Schoenberger, S. 1992. PhD thesis, University of California, Los Angeles.

354. Sengupta, T. K., J. Gordon, and E. K. Spicer. 2001. RegA proteins from phage T4 and RB69 have conserved helix-loop groove RNA binding motifs but different RNA binding specificities. Nucleic Acids Res. 29:1175–1184.

355. Severinova, E., K. Severinov, and S. A. Darst. 1998. Inhibition of *Escherichia coli* RNA polymerase by bacteriophage T4 AsiA. J. Mol. Biol. 279:9–18.

356. Severinova, E., K. Severinov, D. Fenyo, M. Marr, E. N. Brody, J. W. Roberts, B. T. Chait, and S. A. Darst. 1996. Domain organization of the *Escherichia coli* RNA polymerase sigma 70 subunit. J. Mol. Biol. 263: 637–647.

357. Shamoo, Y., A. M. Friedman, M. R. Parsons, W. H. Konigsberg, and T. A. Steitz. 1995. Crystal structure of a replication fork single-stranded DNA binding protein (T4 gp32) complexed to DNA. Nature 376:362–366.

358. Shamoo, Y., A. M. Friedman, M. R. Parsons, W. H. Konigsberg, and T. A. Steitz. 1995. Crystal structure of a replication fork single-stranded DNA binding protein (T4 gp32) complexed to DNA [Erratum]. Nature 376:616.

359. Shamoo, Y., and T. A. Steitz. 1999. Building a replisome from interacting pieces: sliding clamp complexed to a peptide from DNA polymerase and a polymerase editing complex. Cell 99:155–166.

360. Shamoo, Y., A. Tam, W. H. Konigsberg, and K. R. Williams. 1993. Translational repression by the bacteriophage T4 gene *32* protein involves specific recognition of an RNA pseudoknot structure. J. Mol. Biol. 232:89–104.

361. Shamoo, Y., K. R. Webster, K. R. Williams, and W. H. Konigsberg. 1991. A retrovirus-like zinc domain is essential for translational repression of bacteriophage T4 gene 32. J. Biol. Chem. 266:7967–7970.

362. Shamoo, Y., K. R. Williams, and W. H. Konigsberg. 1994. The function of zinc(II) in gene 32 protein (gp32), pp. 305–306. *In* J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, C. Carlson, and E. S. Miller (eds.) Molecular Biology of Bacteriophage T4. American Society for Microbiology, Washington, D.C.

363. Sharma, M., R. L. Ellis, and D. M. Hinton. 1992. Identification of a family of bacteriophage T4 genes encoding proteins similar to those present in group I introns of fungi and phage. Proc. Natl. Acad. Sci. USA 89: 6658–6662.

364. Sharma, M., and D. M. Hinton. 1994. Purification and characterization of the SegA protein of bacteriophage T4, an endonuclease related to proteins encoded by group I introns. J. Bacteriol. 176:6439–6448.

365. Shcherbakov, V., I. Granovsky, L. Plugina, T. Shcherbakova, S. Sizova, K. Pyatkov, M. Shlyapnikov, and O. Shubina. 2002. Focused genetic recombination of bacteriophage T4 initiated by double-strand breaks. Genetics 162: 543–556.

366. Shcherbakov, V. P., S. M. Grebenshchikova, and L. A. Plugina. 1994. [Mechanism of correcting mismatched segments in recombination of bacteriophage T4]. Dokl. Akad. Nauk. 334:541–542.

367. Shcherbakov, V. P., L. A. Plugina, and E. A. Kudryashova. 1995. Marker-dependent recombination in T4 bacteriophage. IV. Recombinational effects of antimutator T4 DNA polymerase. Genetics 140:13–25.

368. Showe, M. K., and L. Onorato. 1978. Kinetic factors and form determination of the head of bacteriophage T4. Proc. Natl. Acad. Sci. USA 75:4165–4169.

369. Shub, D. A., T. Coetzee, D. W. Hall, and M. Belfort. 1994. The self-splicing introns of bacteriophage T4, pp. 186–192. *In* J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, C. Carlson, and E. S. Miller (eds.) Molecular Biology of Bacteriophage T4. American Society for Microbiology, Washington, D.C.

370. Silverstein, J. L., and E. B. Goldberg. 1976. T4 DNA injection. II. Protection of entering DNA from host exonuclease V. Virology 72:212–223.

371. Singer, B. S., L. Gold, P. Gauss, and D. H. Doherty. 1982. Determination of the amount of homology required for recombination in bacteriophage T4. Cell 31:25–33.

372. Sjoberg, B. M., S. Hahne, C. Z. Mathews, C. K. Mathews, K. N. Rand, and M. J. Gait. 1986. The bacteriophage T4 gene for the small subunit of ribonucleotide reductase contains an intron. EMBO J. 5:2031–2036.

373. Smith, D. E., S. J. Tans, S. B. Smith, S. Grimes, D. L. Anderson, and C. Bustamante. 2001. The bacteriophage straight phi29 portal motor can package DNA against a large internal force. Nature 413:748–752.

374. Smith, P. R., and U. Aebi. 1976. Studies of the structure of the T4 bacteriophage tail sheath. I. The recovery of three-dimensional structural information from the extended sheath. J. Mol. Biol. 106:243–271.

375. Snustad, D. P., C. J. Bursch, K. A. Parson, and S. H. Hefeneider. 1976. Mutants of bacteriophage T4 deficient in the ability to induce nuclear disruption: shutoff of host DNA and protein synthesis gene dosage experiments, identification of a restrictive host, and possible biological significance. J. Virol. 18:268–288.

376. Snyder, L., L. Gold, and E. Kutter. 1976. A gene of bacteriophage T4 whose product prevents true late transcription on cytosine-containing T4 DNA. Proc. Natl. Acad. Sci. USA 73:3098–3102.

377. Snyder, L., and G. Kaufmann. 1994. T4 phage exclusion mechanisms, pp. 391–396. *In* J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, C. Carlson, and E. S. Miller (eds.) Molecular Biology of Bacteriophage T4. American Society for Microbiology, Washington, D.C.

378. Solaro, P. C., K. Birkenkamp, P. Pfeiffer, and B. Kemper. 1993. Endonuclease VII of phage T4 triggers mismatch correction in vitro. J. Mol. Biol. 230:868–877.

379. Sozhamannan, S., and B. L. Stitt. 1997. Effects on mRNA degradation by *Escherichia coli* transcription termination factor Rho and pBR322 copy number control protein Rop. J. Mol. Biol. 268:689–703.

380. Stahl, F. W., J. M. Crasemann, C. Yegian, M. M. Stahl, and A. Nakata. 1970. Co-transcribed cistrons in bacteriophage T4. Genetics 64:157–170.

381. Stahl, M. M., L. Thomason, A. R. Poteete, T. Tarkowski, A. Kuzminov, and F. W. Stahl. 1997. Annealing vs. invasion in phage lambda recombination. Genetics 147:961–977.

382. Stevens, A. 1976. A salt-promoted inhibitor of RNA polymerase isolated from T4 phage-infected *E. coli*. *In* R. Losick and M. Chamberlin (eds.) RNA Polymerase. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

383. Stitt, B., and D. Hinton. 1994. Regulation of middle-mode transcription, pp. 142–160. *In* J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, C. Carlson, and E. S. Miller (eds.) Molecular Biology of Bacteriophage T4. American Society for Microbiology, Washington, D.C.

384. Stitt, B. L., and E. S. Kempner. 1996. Structure–function relationships in *Escherichia coli* transcription termination protein Rho revealed by radiation target analysis. Arch. Biochem. Biophys. 334:268–276.

385. Stitt, B. L., and G. Mosig. 1989. Impaired expression of certain prereplicative bacteriophage T4 genes explains impaired T4 DNA synthesis in *Escherichia coli* rho (nusD) mutants. J. Bacteriol. 171:3872–3880.

386. Stohr, B. A., and K. N. Kreuzer. 2002. Coordination of DNA ends during double-strand-break repair in bacteriophage T4. Genetics 162:1019–1030.

387. Stohr, B. A., and K. N. Kreuzer. 2001. Repair of topoisomerase-mediated DNA damage in bacteriophage T4. Genetics 158:19–28.

388. Streisinger, G. 1966. Terminal redundancy, or all's well that ends well, pp. 335–340. *In* J. Cairns, G. S. Stent, and J. D. Watson (eds.) Phage and the Origins of Molecular Biology. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

389. Streisinger, G., R. S. Edgar, and G. H. Denhardt. 1964. Chromosome structure in phage T4. I. Circularity of the linkage map. Proc. Natl. Acad. Sci. USA 51:775–779.

390. Streisinger, G., Y. Okada, J. Emrich, J. Newton, A. Tsugita, E. Terzaghi, and M. Inouye. 1966. Frameshift mutations and the genetic code. Cold Spring Harb. Symp. Quant. Biol. 31:77–84.

391. Symmons, M. F., M. G. Williams, B. F. Luisi, G. H. Jones, and A. J. Carpousis. 2002. Running rings around RNA: a superfamily of phosphate-dependent RNases. Trends Biochem. Sci. 27:11–18.

392. Symonds, N., H. Heindl, and P. White. 1973. Radiation sensitive mutants of phage T4. A comparative study. Mol. Gen. Genet. 120:253–259.

393. Szostak, J. W., T. L. Orr-Weaver, R. J. Rothstein, and F. W. Stahl. 1983. The double-strand-break repair model for recombination. Cell 33:25–35.

394. Tao, Y., S. V. Strelkov, V. V. Mesyanzhinov, and M. G. Rossmann. 1997. Structure of bacteriophage T4 fibritin: a segmented coiled coil and the role of the C-terminal domain. Structure 5:789–798.

395. Tarahovsky, Y. S., A. A. Khusainov, A. A. Deev, and Y. V. Kim. 1991. Membrane fusion during infection of *Escherichia coli* cells by phage T4. FEBS Lett. 289:18–22.

396. Tetart, F., C. Desplats, and H. M. Krisch. 1998. Genome plasticity in the distal tail fiber locus of the T-even bacteriophage: recombination between conserved motifs swaps adhesin specificity. J. Mol. Biol. 282: 543–556.

397. Tetart, F., C. Desplats, M. Kutateladze, C. Monod, H. W. Ackermann, and H. M. Krisch. 2001. Phylogeny of the major head and tail genes of the wide-ranging T4-type bacteriophages. J. Bacteriol. 183:358–366.

398. Tetart, F., F. Repoila, C. Monod, and H. M. Krisch. 1996. Bacteriophage T4 host range is expanded by duplications of a small domain of the tail fiber adhesin. J. Mol. Biol. 258:726–731.

399. Thomas, C. A., and L. A. MacHattie. 1967. The anatomy of viral DNA molecules. Ann. Rev. Biochem. 36:485–518.

400. Thomas, E., F. Zucker, and E. Kutter. 2003. A phylogenetic analysis of selected T4 proteins. Manuscript in preparation.

401. Thompson, R. J., J. P. Davies, G. Lin, and G. Mosig (eds.). 1990. Modulation of transcription by altered torsional stress, upstream silencers, and DNA-binding proteins, pp. 227–240 In K. Drlica and M. Riley (eds.) The Bacterial Chromosome. American Society for Microbiology, Washington, D.C.

402. Tinker, R. L., G. M. Sanders, K. Severinov, G. A. Kassavetis, and E. P. Geiduschek. 1995. The COOH-terminal domain of the RNA polymerase alpha subunit in transcriptional enhancement and deactivation at the bacteriophage T4 late promoter. J. Biol. Chem. 270: 15899–15907.

403. Torgov, M. Y., D. M. Janzen, and M. K. Reddy. 1998. Efficiency and frequency of translational coupling between the bacteriophage T4 clamp loader genes. J. Bacteriol. 180:4339–4343.

404. Trojanowska, M., E. S. Miller, J. Karam, G. Stormo, and L. Gold. 1984. The bacteriophage T4 *regA* gene: primary sequence of a translational repressor. Nucleic Acids Res. 12:5979–5993.
405. Unnithan, S., L. Green, L. Morrissey, J. Binkley, B. Singer, J. Karam, and L. Gold. 1990. Binding of the bacteriophage T4 regA protein to mRNA targets: an initiator AUG is required. Nucleic Acids Res. 18:7083–7092.
406. Uzan, M., R. Favre, and E. Brody. 1988. A nuclease that cuts specifically in the ribosome binding site of some T4 mRNAs. Proc. Natl. Acad. Sci. USA 85:8895–8899.
407. Vaiskunaite, R., A. Miller, L. Davenport, and G. Mosig. 1999. Two new early bacteriophage T4 genes, *repEA* and *repEB*, that are important for DNA replication initiated from origin E. J. Bacteriol. 181:7115–7125.
408. van Raaij, M. J., G. Schoehn, M. R. Burda, and S. Miller. 2001. Crystal structure of a heat- and protease-stable part of the bacteriophage T4 short tail fibre. J. Mol. Biol. 314:1137–1146.
409. van Raaij, M. J., G. Schoehn, M. Jaquinod, K. Ashman, M. R. Burda, and S. Miller. 2001. Identification and crystallisation of a heat- and protease-stable fragment of the bacteriophage T4 short tail fibre. Biol. Chem. 382: 1049–1055.
410. Vanzo, N. F., Y. S. Li, B. Py, E. Blum, C. F. Higgins, L. C. Raynal, H. M. Krisch, and A. J. Carpousis. 1998. Ribonuclease E organizes the protein interactions in the *Escherichia coli* RNA degradosome. Genes Dev. 12:2770–2781.
411. Villarreal, L. P., and V. R. DeFilippis. 2000. A hypothesis for DNA viruses as the origin of eukaryotic replication proteins. J. Virol. 74:7079–7084.
412. Villemain, J. L., Y. Ma, D. P. Giedroc, and S. W. Morrical. 2000. Mutations in the N-terminal cooperativity domain of gene *32* protein alter properties of the T4 DNA replication and recombination systems. J. Biol. Chem. 275: 31496–31504.
413. Visconti, N., and M. Delbrück. 1953. The mechanism of genetic recombination in phage. Genetics 38:5–33.
414. Volkin, E., and L. Astrachan. 1956. Phosphorus incorporation in *Escherichia coli* ribo-nucleic acid after infection with bacteriophage T2. Virology 2:149–161.
415. von Hippel, P. H., S. C. Kowalczykowski, N. Lonberg, J. W. Newport, L. S. Paul, G. D. Stormo, and L. Gold. 1983. Autoregulation of expression of T4 gene 32: a quantitative analysis, pp. 202–207. *In* C. K. Mathews, E. Kutter, G. Mosig, and P. B. Berget (eds.) Bacteriophage T4. American Society for Microbiology, Washington, D.C.
416. Wais, A. C., and E. B. Goldberg. 1969. Growth and transformation of phage T4 in *Escherichia coli* B-4, *Salmonella*, *Aerobacter*, *Proteus*, and *Serratia*. Virology 39:153–161.
417. Wakem, L. P., and K. Ebisuzaki. 1984. An analysis of DNA repair and recombination functions of bacteriophage T4 by means of suppressors: the role of das. Virology 137:324–330.
418. Wang, C. C., A. Pavlov, and J. D. Karam. 1997. Evolution of RNA-binding specificity in T4 DNA polymerase. J. Biol. Chem. 272:17703–17710.
419. Wang, G. R., A. Vianelli, and E. B. Goldberg. 2000. Bacteriophage T4 self-assembly: in vitro reconstitution of recombinant gp2 into infectious phage. J. Bacteriol. 182:672–679.
420. Wang, J., A. K. Sattar, C. C. Wang, J. D. Karam, W. H. Konigsberg, and T. A. Steitz. 1997. Crystal structure of a pol alpha family replication DNA polymerase from bacteriophage RB69. Cell 89:1087–1099.
421. Wang, J., P. Yu, T. C. Lin, W. H. Konigsberg, and T. A. Steitz. 1996. Crystal structures of an NH_2-terminal fragment of T4 DNA polymerase and its complexes with single-stranded DNA and with divalent metal ions. Biochemistry 35:8110–8119.
422. Washburn, R. S., and B. L. Stitt. 1996. In vitro characterization of transcription termination factor Rho from *Escherichia coli* rho(nusD) mutants. J. Mol. Biol. 260:332–346.
423. Weaver, L. H., D. Rennell, A. R. Poteete, and B. W. Mathews. 1985. Structure of phage P22 gene *19* lysozyme inferred from its homology with phage T4 lysozyme. Implications for lysozyme evolution. J. Mol. Biol. 184:739–741.
424. Webster, K. R., and E. K. Spicer. 1990. Characterization of bacteriophage T4 regA protein–nucleic acid interactions. J. Biol. Chem. 265:19007–19014.
425. Werner, R. 1968. Distribution of growing points in DNA of bacteriophage T4. J. Mol. Biol. 33:679–692.
426. Werner, R. 1969. Initiation and propagation of growing points on the DNA of phage T4. Cold Spring Harbor Symp. Quant. Biol. 33:501–507.
427. Wheeler, L. J., N. B. Ray, C. Ungermann, S. P. Hendricks, M. A. Bernard, E. S. Hanson, and C. K. Mathews. 1996. T4 phage gene *32* protein as a candidate organizing factor for the deoxyribonucleoside triphosphate synthetase complex. J. Biol. Chem. 271:11156–11162.
428. Wiberg, J. S., and J. D. Karam. 1983. Translational regulation in T4 phage development, pp. 193–201. *In* C. K. Mathews, E. Kutter, G. Mosig, and P. B. Berget (eds.) Bacteriophage T4. American Society for Microbiology, Washington, D.C.
429. Wiberg, J. S., S. Mendelsohn, V. Warner, K. Hercules, C. Aldrich, and J. L. Munro. 1973. SP62, a viable mutant of bacteriophage T4D defective in regulation of phage enzyme synthesis. J. Virol. 12:775–792.
430. Wilkens, K., and W. Ruger. 1996. Characterization of bacteriophage T4 early promoters in vivo with a new promoter probe vector. Plasmid 35:108–120.
431. Wilkens, K., B. Tiemann, F. Bazan, and W. Ruger. 1997. ADP-ribosylation and early transcription regulation by bacteriophage T4, pp. 71–82. *In* F. Haag and F. Koch-Nolte (eds.) ADP-Ribosylation in Animal Tissue. Plenum Press, New York.
432. Williams, K. P., G. A. Kassavetis, D. R. Herendeen, and E. P. Geiduschek. 1994. Regulation of late-gene expression, pp. 161–175. *In* J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, C. Carlson, and E. S. Miller (eds.) Molecular Biology of Bacteriophage T4. American Society for Microbiology, Washington, D.C.
433. Williams, K. R., Y. Shamoo, E. K. Spicer, J. E. Coleman, and W. H. Konigsberg. 1994. Correlating structure to function in proteins: T4 gp32 as a prototype, pp. 301–304.

In J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, C. Carlson, and E. S. Miller (eds.) Molecular Biology of Bacteriophage T4. American Society for Microbiology, Washington, D.C.

434. Winter, R. B., L. Morrissey, P. Gauss, L. Gold, T. Hsu, and J. Karam. 1987. Bacteriophage T4 regA protein binds to mRNAs and prevents translation initiation. Proc. Natl. Acad. Sci. USA 84:7822–7826.
435. Wood, W. B., and M. P. Conley. 1979. Attachment of tail fibers in bacteriophage T4 assembly: role of the phage whiskers. J. Mol. Biol. 127:15–29.
436. Wood, W. B., F. A. Eiserling, and R. A. Crowther. 1994. Long tail fibers: genes, proteins, structure and assembly, pp. 282–290. *In* J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, C. Carlson, and E. S. Miller (eds.) Molecular Biology of Bacteriophage T4. American Society for Microbiology, Washington, D.C.
437. Wu, C. H., and L. W. Black. 1995. Mutational analysis of the sequence-specific recombination box for amplification of gene *17* of bacteriophage T4. J. Mol. Biol. 247: 604–617.
438. Wu, C. H., H. Lin, and L. W. Black. 1995. Bacteriophage T4 gene *17* amplification mutants: evidence for initiation by the T4 terminase subunit gp16. J. Mol. Biol. 247:523–528.
439. Yee, J. K., and R. C. Marsh. 1985. Locations of bacteriophage T4 origins of replication. J. Virol. 54:271–277.
440. Yeh, L. S., T. Hsu, and J. D. Karam. 1998. Divergence of a DNA replication gene cluster in the T4-related bacteriophage RB69. J. Bacteriol. 180:2005–2013.
441. Youil, R., B. W. Kemper, and R. G. Cotton. 1995. Screening for mutations by enzyme mismatch cleavage with T4 endonuclease VII. Proc. Natl. Acad. Sci. USA 92:87–91.
442. Yu, Y. T., and L. Snyder. 1994. Translation elongation factor Tu cleaved by a phage-exclusion system. Proc. Natl. Acad. Sci. USA 91:802–806.

19

Bacteriophage T5

JON R. SAYERS

Several features of phage type 5 make this a very interesting and rather unusual lytic bacteriophage. Phage T5 does not encode an RNA polymerase, unlike the T7 phage, nor does it contain unusual DNA modifications such as the 5-hydroxymethyl deoxycytidine present in phage T4. T5 DNA carries some of the strongest known prokaryotic promoters but they are regulated in a strictly controlled temporal sequence. Perhaps the most unusual feature is that the phage's DNA enters the host in a two-step transfer mechanism. Transfer of the viral genome from the virion to host pauses after injection of the first 10 kb of the 120 kbp linear genome, until the pre-early phage-encoded proteins have been produced. The DNA itself is unusual in that it contains a number of cryptic nicks in one strand of its double-stranded genome. The phage is able to destroy host cell DNA rapidly, yet can quickly inactivate the nuclease responsible for this degradation.

The T5 group includes phages BF23, 29α, BG3, and PB (52). Of these only T5 and BF23 have been studied in any detail. However, in comparison with phages λ, T4, and T7, relatively little work has been published on these fascinating phages. This chapter will present a general overview of what is known about their life cycles and will concentrate particularly on advances made since the last major review of the group was published (56).

Physical Properties and Attachment to the Host

Bacteriophage T5 belongs to Bradley's morphological group B (12). A long, noncontractile tail of 200 nm by 12 nm (11) is attached to one of the vertices of the icosahedral head (2). Bacteriophage T5 DNA consists of a linear double-stranded DNA genome approximately 121,300 bp long. The phage genome possesses terminal redundancies in the form of 10,160 bp direct repeats (65, 66). These repeats carry the so-called pre-early genes (figure 19-1). An unusual feature of T5 DNA is that one strand is nicked in several places (1). The role and origins of these nicks remain cryptic despite early reports of four nucleases able to introduce such interruptions into the DNA backbone (69). However, the termini of several nicks have been analyzed and found to contain the consensus sequence–NPurine/5′pGCGCN–(63).

Irreversible adsorption of the virion to the host follows once the product of the T5 *oad* gene (or *hrs* in BF23) binds to the cellular receptor (34). The protein encoded by *oad*, pb5, binds to the FhuA protein (32, 59). The orthologous gene in BF23, *hrs*, encodes a related but divergent receptor-binding protein (44, 58). Sequence analysis of the T5 and BF23 receptor-binding proteins reveals that they share some conserved motifs (58). To date this is the only major variation observed between T5 and BF23 sequence data. The BF23 pb5 protein binds to BtuB, the *E. coli* vitamin B_{12} outer membrane transport protein (10). The T5 receptor, FhuA, is a multifunctional membrane protein involved in the uptake of ferrichrome-iron (34).

Elegant work by Killman and coworkers has identified a short peptide sequence able to trigger complete release of phage T5 DNA from the virion (43). This peptide sequence, APADKGHY, maps to a loop on the extracellular side of the β-barrel membrane-protein FhuA, the T5 receptor. The structure of this monomeric membrane protein has been determined by Ferguson et al. (21) and is shown in figure 19-2. The protein has two domains: a β-barrel composed of 22 antiparallel β-strands and a structurally distinct "cork" domain. The cork, comprised of four-stranded β-sheets and four short α-helices, is inside the β-barrel and thus blocks the channel. Binding of T5 virions to FhuA triggers a conformational change thereby opening a channel large enough to allow an efflux of Fe(III)-ferrichrome and the passage of T5 DNA (51).

First-Step Transfer of DNA

Once irreversibly bound to FhuA, the phage DNA is transferred into the host cell in a two-step process. The structure of the FhuA protein has been determined (64). It consists of a β-barrel formed by the C-terminal 556 amino

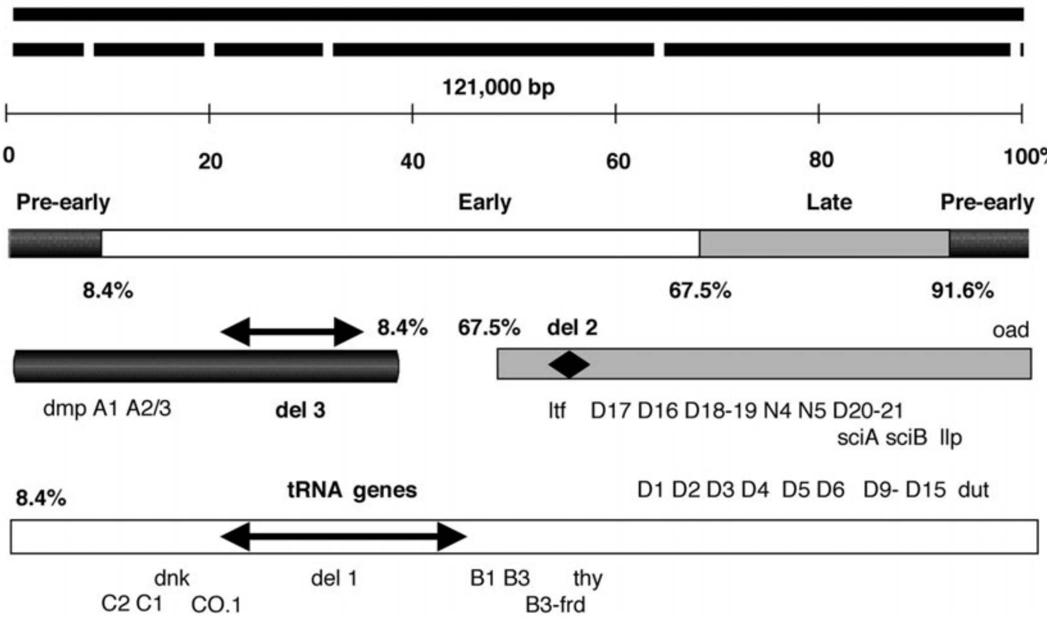

121,000 bp
0
20
40
60
80
100%
Pre-early
Early
Late
Pre-early
8.4%
67.5%
91.6%
8.4%
67.5%
del 2
oad
dmp A1 A2/3
del 3
ltf D17 D16 D18-19 N4 N5 D20-21
sciA sciB llp
8.4%
tRNA genes
D1 D2 D3 D4 D5 D6 D9- D15 dut
dnk
del 1
B1 B3
thy
C2 C1
CO.1
B3-frd

Figure 19-1 Diagram relating the phage T5 physical and genetic map. The double-stranded DNA duplex (black lines) consists of one continuous (upper) and one nicked strand. Pre-early genes are encoded on the terminal repeats (dark shaded). The early and late regions are shown in white and grey, respectively. Genes coded on the top and bottom strands are shown above and below the bars respectively. Black arrows and a diamond show the positions of viable deletions.

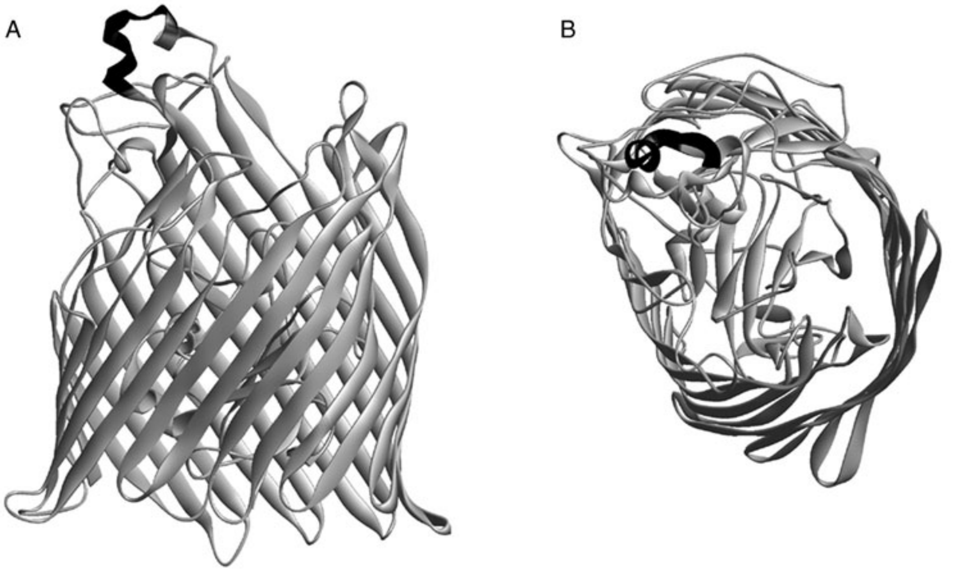


Figure 19-2 The T5 receptor protein. A: Structure of the T5 receptor, FhuA, as determined by Ferguson viewed in the plane of the membrane (21). The β-barrel traverses the outer membrane. Note the black loop that interacts with the T5 receptor binding protein pb5 (43). There is also a secondary interaction with a region within the β-barrel. B: A ‘‘phage's eye view'' of the receptor and the cork domain which blocks the channel until the phage binds.

acid residues. These residues form a channel or pore-like structure blocked with a "plug" composed of amino acid residues 20–157. Presumably, once irreversibly bound to pb5, the FhuA or BtuB transmembrane proteins undergo some form of conformational change to allow injection of the nucleic acid. It is the left end of the genome which is transferred into the host cell first (35, 36, 45, 74). This process has been studied in vitro using FhuA protein incorporated into unilamellar vesicles made of natural phospholipids (46). Under these conditions the entire T5 genome is rapidly transferred into the vesicle, where the DNA remains in a densely packed form. This is in stark contrast to the situation in vivo where DNA is transferred in a two-step process. Upon irreversible binding to the cell surface's FhuA or BtuB protein, the process of first-step transfer (FST) occurs during the first 2 minutes, in which approximately the first 8% of the left-hand end of the linear duplex genome is injected into the host cell (48). There then follows a pause of about 5 minutes during which time T5-specific messenger RNA (mRNA) is synthesized and translated into the pre-early proteins. These proteins, A1 and A2–3, must be produced in order for the remaining T5 DNA to enter the cell in a process known as second-step transfer (SST) (47). It has been shown that this process requires the presence of calcium ions at approximately 1 mM. Calcium is required to maintain polarization of the bacterial cytoplasmic membrane, supporting the notion that the pores opened during FST must be closed in order for the infection to proceed (7). If calcium is not present, few or no proteins are transcribed from the FST DNA and the infection aborts unless calcium is added back to the medium (8).

Pre-Early Proteins

Bacteriophage T5 contains some of the strongest promoters yet characterized, with high affinity for the host cell RNA polymerase (27, 84). The FST DNA contains at least three promoters which drive the production of pre-early mRNA immediately upon entry (83). Biochemical studies suggest that up to 10 pre-early proteins are present in phage-infected cells but not all are essential, since a viable mutant of BF23, which carries a deletion (del1), fails to synthesize three or four pre-early protiens (55). The best characterized pre-early genes are A1, A2–3, and *dmp*. Expression of the pre-early genes has dire consequences for the host cell: synthesis of *E. coli* DNA, RNA, and proteins are all rapidly shut-off; most of the host cell DNA is degraded (49, 88); host enzymes such as *Eco*RI, *recBC*, DNA methylase, and uracil-DNA glycosylase are all inactivated (56). The identity of the nuclease responsible for the rapid destruction of host cell DNA is not yet clear, but A1 mutants fail to degrade DNA, leading to speculation that this is the nuclease responsible.

The A1 protein has been localized to both inner and outer membranes. It has a molecular weight of approximately 57 kDa, and partial sequencing reveals that the N-terminal region possesses a signal peptide leader sequence which is consistent with membrane association (86). A1 appears to form multimers with itself and hetero-oligomers with the smaller A2–3 protein (4). The A1 homo-oligomer is approximately 244 kDa, while the hetero-oligomeric complex with the A2 polypeptide is larger, at around 364 kDa (4). Sequencing of part of the BF23 pre-early region showed genes A2 and A3 to be one gene, renamed A2–3, which encodes a protein of 13.8 kDa (70, 86). The BF23 and T5 A2–3 proteins share over 95% identity and show some similarity to a number of nucleotide binding proteins and a lipoprotein (23, 79). The A2–3 protein is implicated in altering the host cell membrane structure (78) and the purified protein is able to bind DNA in vitro (80).

Another unusual feature of T5 infection is the excretion of bases and nucleotides from the cell during the first few minutes of infection. Host cell DNA degradation products (deoxyribonucleoside 5′ monophosphates, dNMPs) are converted to deoxyribonucleosides by the pre-early product of *dmp*, a 5′ deoxyribonucleotidase, which are further broken down to bases and excreted by cellular enzymes (85). Presumably this excretion is required to avoid the potentially damaging build-up of dNMPs that would otherwise occur (61). The nucleotidase activity decays rapidly after SST has taken place, implying that an early gene is responsible for shutting down this enzyme (5).

Pausing and Second-Step Transfer

What causes the transfer of DNA to pause at the FST stage? It had been proposed that a complex secondary structure or physical discontinuity on the DNA might provide the injection stop signal (ISS) (75). Parts of the FST region, including the ISS, have been sequenced. This region does indeed contain sequences with a high potential for forming complex secondary structures, such as multiple direct repeats of up to 18 bp, inverted repeats and an 18 bp palindrome. These repeats include three 31 bp repeat units contained on a 99 bp sequence that could form mutually exclusive stem-and-loop structures. Another pair of 21 bp repeats contain two sequences resembling DnaA protein-binding sites (37). However, given that exposure of T5 bacteriophage to purified FhuA alone is enough to cause complete ejection of the entire genome from the head, it is clear that some other interactions must be involved to halt FST at the ISS.

The presence of elaborate secondary structures and potential DnaA-binding sites at the ISS suggests the involvement of host DNA-binding proteins (37). In addition, one of the cryptic nicks occurs at the base of a 9 bp palindrome in this region. It seems likely that host cell molecules, most probably DNA-binding proteins, bind the ISS, halting transfer at FST. Given that the A1 protein can form oligomers with A2–3, which in turn has DNA-binding activity, it seems likely that these large protein complexes bind to the elaborate structures potentially formed at the ISS. Perhaps these phage-encoded proteins act to initiate SST by displacing host proteins bound at the ISS, subsequently allowing SST to begin. Certainly mutants defective in A2–3 are unable to complete SST, but do degrade host cell DNA, and SST can only occur after synthesis of A1 and A2–3 polypeptides (47).

Early Genes and Their Roles

Many of the devastating effectors of pre-early gene expression become inactivated before or shortly after SST takes place. The potent nuclease activity responsible for destruction of the host's DNA is neutralized and the strong pre-early promoters are silenced, probably due to interactions of the host cell RNA polymerase with the A1 protein and another pre-early gene product (53). The early region also contains an as yet unidentified factor required to shut off the phage-specified nucleotidase activity (5). This shutoff is necessary for efficient replication to occur. Another early gene, *llp*, encodes a lipoprotein (18) whose function is to block superinfection of the host. This is accomplished in a process designated as lytic conversion (64) in which the lipoprotein interacts directly with the FhuA receptor. This process requires a large excess of the lipoprotein, at least in vitro, where the ratio of this protein to receptor needs to be in excess of 10:1 in order to fully block the interaction of T5 virions with FhuA (64).

The SST DNA contains a large deletable section in the early gene region (del 2, 21.1–32.3% on the map) encoding 24 tRNAs (76). These genes are presumably required to facilitate efficient translation of phage mRNA required for viral replication. Host tRNA, produced prior to the extensive degradation of host DNA shortly after FST, can suffice to support modest replication rates upon infection with T5st(0) (a phage in which much of the tRNA encoding region has been deleted), but replication times become extended. A structural comparison of phage-encoded tRNAs suggested that they may possess some unusual features. For example, the tRNALeu and tRNATrp have a longer anticodon loop than the corresponding bacterial tRNAs and could act as frameshift mutation suppressors (76). This region also contains several open reading frames of unknown function.

The SST DNA contains late, early, and pre-early promoters, the latter being on the second terminal repeat. However, expression of genes on the SST DNA is limited to the early genes for 8 minutes after SST. Only after certain early gene products have accumulated can the late genes become active, which occurs at about 12 minutes after infection. Early genes C2, D5, and possibly D15 are thought to be involved in regulation of T5 gene expression. The role of the D15 nuclease will be discussed later. The 90 kDa protein encoded by C2 appears to be weakly associated with host RNAP and could act as an alternative σ factor, thereby shifting specificity from early to late promoters (16, 81, 82). The D5 gene product appears to have a more complex role, being involved in both positive and negative regulation as well as in DNA replication (16, 54). Biochemical studies on the purified D5 protein suggest that it is a highly soluble 29 kDa protein, able to bind both double- and single-stranded DNA in vitro (68) and is associated with the T5 transcription:replication complex (22).

A large region of the SST DNA is devoted to encoding genes involved in DNA replication. Cells infected with T5 produce more DNA packaged as phage than was originally present in the cell prior to infection. This large synthetic burden is carried out with the aid of many phage-specified proteins. These include DNA polymerase (the D9 gene product, formerly D7–8-9), 5′–3′ exonuclease (D15), helicase (D10), DNA ligase, dihydrofolate reductase (*dhr*), ribonucleotide reductase (possibly B1 or B2), deoxynucleoside monophosphokinase (*dnk*, C0.3) (57), thioredoxin and thymidylate synthetase (reviewed in 52). The *dnk* gene product has recently been purified from phage-infected cells and its biological activity and N-terminal sequence determined (19).

The D15 exonuclease gene was sequenced by Kaliman et al. (40) and shown to encode a protein of 291 amino acids. A consensus *E. coli* promoter is located upstream of the D15 Shine–Dalgarno region. Overexpression of the enzyme has allowed the biochemistry and biophysics of this protein to be investigated (19, 72). The D15 nuclease functions as an exonuclease in the 5′–3′ manner, releasing mono-, or short polynucleotides depending on the substrate. In addition to this activity, the enzyme is an efficient flap endonuclease, able to cleave bifurcated DNA possessing a free 5′ arm (13). The D15 nuclease also possesses nanomolar binding affinities for DNA substrates with single-stranded arms (26). The crystal structure of this phage enzyme revealed the presence of a helical arch or clamp, thought to be involved in DNA binding or threading of substrates (13). Sequence analysis shows that this enzyme is highly homologous with the small fragment of DNA polymerase I, that is, amino acids 1–323 carrying the 5′–3′ exonuclease activity of the polymerase holoenzyme. So far this is the only T5 protein whose crystal structure has been determined (figure 19-3).

It has been suggested that D15 may play a role in processing T5 DNA containing pre-existing nicks by cleaving the intact strand opposite the nick (60). Biochemical studies on the overexpressed enzyme do suggest that, in addition to flap endonuclease activity, the enzyme is able to cleave in a purely endonucleolytic manner in a reaction requiring a nicked substrate (73). The same study provided some support for early reports suggesting that the D15 protein may be involved in switching on late genes (15). Although little further work has been reported in this area, the D15 nuclease is able to introduce gaps into nicked substrates in vitro. It has been suggested that single-stranded regions in replicating T5 DNA are needed for late gene expression (15).

The D9 DNA polymerase has been isolated in the T5 transcription-replication complex together with the D5 and D15 polypeptides (22). The D15 nuclease is clearly involved in DNA replication, as had been long suspected (24, 25). Bacteriophage T4 contains a close homolog of T5 D15 known as T4 RNaseH (38). The structures of these

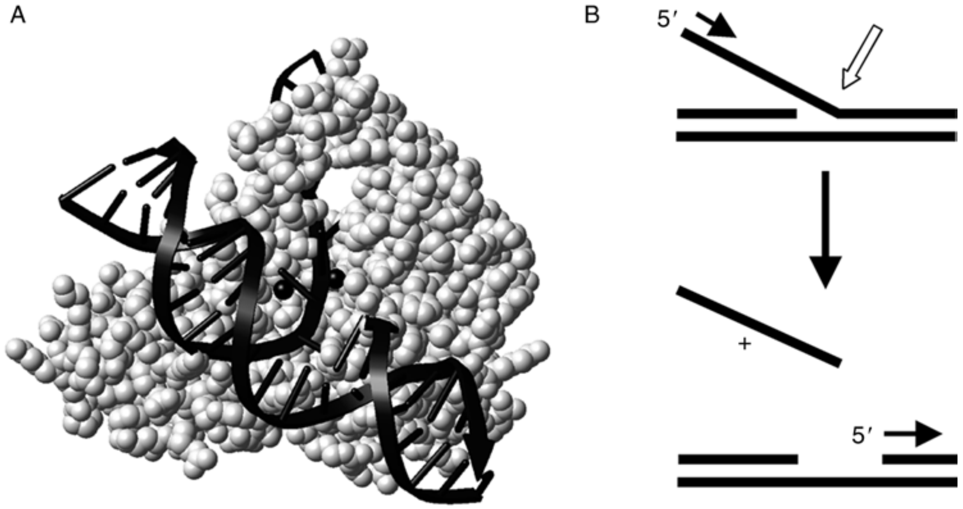


Figure 19-3 The phage T5 D15 nuclease. A: Crystal structure of the D15 exonuclease or flap endonuclease. A conceptual model of a bound DNA flap substrate is shown (13, 19). B: Endo-nucleolytic (open arrow) and exo-nucleolytic modes (black arrow) of DNA cleavage and gapping reactions at a nick.

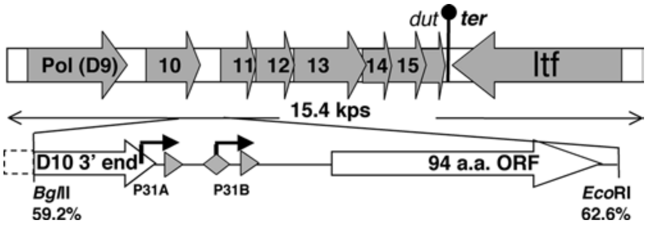


Figure 19-4 Diagrammatic representation of a 15,000 bp region covering the junction between early (D genes) and late genes. This sequence was assembled from three sequences, accession numbers M64047, M24354, and AY543070. The position of a strong terminator of transcription (*ter*) is indicated. The D9 DNA polymerase is preceded by a promoter (not shown). D10 is a putative helicase, D11 and D14 have no obvious functions or homologs of known function. Genes D12 and D13 are thought to be part of a repair nuclease. The D15 exonuclease gene overlaps with the deoxyuridine triphosphatase (*dut*) gene, both of which are served by a promoter just upstream. The long tail fibre gene (*ltf*) is transcribed in the opposite orientation to the early genes. In addition to the reading frames shown, the fragment contains numerous ORFs with little or no homology to any database deposits. An enlargement of a region defined by *Eco*RI and *Bgl*II restriction sites is shown. Tandem promoters P_{31A} and P_{31B} are separated by 100 bases (39). Their −10 sequences are both overlapped by a 20 bp direct repeat (triangles). P_{31B} possesses an 18 bp imperfect palindrome (diamond) that overlaps with the P_{31B} −35 element. A cryptic 94 amino acid ORF is situated downstream and must be present in order to clone fragments carrying P31B (39).

enzymes are very similar indeed (3). This has been shown to be important for processing of Okazaki fragments from replicating DNA lagging strands (6). These RNA oligomers are removed largely through the action of the 5′–3′ exonucleolytic function of the 5′ nuclease. Together the D9 polymerase, and D15 5′ nuclease proteins make up the equivalent of the *E. coli* DNA polymerase I holoenzyme, possessing 5′–3′ exonuclease (D15 nuclease), polymerase, and proofreading activities (D9 polymerase). The D9 gene has been sequenced and shown to encode a polymerase with associated 3′–5′ proofreading exonuclease activity as expected (50). This enzyme is highly processive and can carry out strand displacement synthesis, but extensive characterization of this enzyme has been limited due to lack of solubility of the over-expressed protein (14).

A large contiguous sequence encompassing the region from the D9 polymerase to the end of one of the late genes (including the *ltf* gene encoding the long tail fiber protein) can be assembled from sequences M64047, M24354, and AY543070 (41, 42, 50), which allows an unambiguous ordering of several genes in this region (figure 19-4). The sequence also reveals the presence of tandem promoters P31A and P31B, which appear to control expression of a putative D12-D13 exonuclease, possibly involved in DNA repair (42). Interestingly, we were able to clone a fragment carrying these tandem promoters provided that an adjacent open reading frame (ORF) was present on the cloned fragment of T5 DNA (see figure 19-4). This was possible despite the observations that the closely related D9 promoter could not be cloned on a multicopy plasmid due to the high strength of that promoter (14). Disruption of the ORF immediately downstream of P31B led to recovery of clones lacking the strong promoter P31B. Loss of this promoter was due to recombination between 20 bp direct repeats flanking P31B, leaving only the weaker P31A promoter (J. Sorrell and J. R. Sayers, unpublished data). No obvious transcription terminator was present on this fragment. Kaliman et al. (39) reported similar observations when working with the same fragment. Stable cloning of strong T5 promoters often requires the presence of a downstream terminator as described by Bujard and coworkers (28). We suggest that the ORF downstream of P31B might encode a repressor capable of inactivating this strong promoter, thereby providing a stabilizing influence on the recombinant plasmid (J. Sorrell and J. R. Sayers, unpublished data).

Replication

The replication of T5 DNA has not been as well studied as the replication of T4 and other phages. Since the last major review of the T5 group (56), little progress has been made in determining the structure of replication intermediates present in T5-infected cells. Replication was shown to proceed bidirectionally from multiple, internal origins. A primary origin of replication is located near the center of the genome. Significant numbers of circular T5 DNA molecules are observed during the later stages of infection. These replicative circular molecules appear to be in either a theta or sigma configuration (9). Replication forks, loops, and circular structures, similar to those found in other large phages, were also observed by Everett (20). Phage capsid structures were associated with both mature phage-length DNA and concatemeric molecules (20). It also appears that host DNA gyrase is required for T5 DNA replication and for late gene expression (17).

Late Genes

The late genes encoding structural proteins reside to the right of the D15 gene and their order has been determined by Heller (32). The virion is composed of 15 or more proteins (32, 87). The most abundant is the major head protein encoded by D20–21 (32 kDa, 730 copies), N4 encoding the major tail protein (58 kDa, 120 copies) and the N5 gene product, producing another head protein (19 kDa, 114 copies). In BF23 a minor tail protein (gene 24) with a mass of 34 kDa has been identified along with gene 25 which encodes a major tail protein of 50 kDa, possibly corresponding to gene N4 in T5 (62). The gene encoding the L-shaped tail-fiber (*ltf*) has been sequenced and shown to direct the synthesis of a protein of approximately 148 kDa, in reasonable agreement with the previously reported mass of 125 kDa (33, 42). Each virion contains six to nine copies of this bent tail fiber, which are thought to mediate reversible attachment to the host cell via interactions with polymannose O-antigens (30, 31). However, the existence of a viable deletion mutant indicates that this gene is not essential (32, 67, 71). The bent tail fiber sequence shares strong sequence similarity with the side tail fiber protein of λ phage and it has been suggested that there is extensive horizontal transfer of such genes between phages (29). In addition to the *oad* and *ltf* genes, the only other sequenced region of the late genes appears to encode a protein with an unexpected and somewhat surprising sequence. A partial sequence from the late region of BF23 was determined and found to contain collagen-like repeats (77). Similar genes are present in other phages, despite the fact that no such collagen-like genes have been identified to date among the eubacteria. This raises the intriguing possibility that the phages have been able to accept gene transfer from the eukaryotes or have evolved the collagen repeat independently. The function of the collagen-like protein remains unknown, but it seems that such a protein is likely to have a structural function, particularly as it appears to be expressed from a late gene.

Concluding Remarks

Many features of phage T5 remain enigmatic. What is the identity of the nuclease responsible for host cell DNA degradation, and what protects the phage DNA from its attentions? What are the molecular mechanisms governing the pause in DNA injection after FST and how is SST initiated? Which protein(s) introduces the DNA nicks and what is their role in vivo? How do phage proteins subvert the host RNA polymerase, changing its specificity in tune with the demands for pre-early, early, and late gene expression? How do phage proteins overcome hostile host functions such as restriction and methylation? Why has such an interesting phage remained so much a mystery? In the past this has been due to the difficulties of cloning large fragments of T5 DNA due to promoter strength and the toxicity of its gene products. It is clear that the technical ability to sequence this phage without cloning exists. It has long been possible to sequence large double-stranded DNA such as λ using the walking-primer approach (63). Two groups have independently sequenced the T5 genome. A Russian group released their sequence in April 2004 (accession number AY543070), and French researchers afterwards shortly (accession number AY692264). The genome consists of 121,750 base pairs and some 162 genes are annotated. Access to the genome sequence will undoubtedly aid and stimulate researchers in the T5 field, but a cursory analysis reveals a large proportion of proteins of unknown function, many of which do not have any reasonably close homolog in the current databases (not even cryptic ORFs from other phages or bacteria). Moreover, T5 possesses many ORFs not identified on the genetic map. Even in 2005, most of the questions about T5 biology, first posed 30 or more years ago, remain unanswered.

References

1. Abelson, J. and C. A. Thomas. 1966. The anatomy of the T5 bacteriophage DNA molecule. J. Mol. Biol. 18:262.
2. Anderson, T. F. 1973. Morphologies of bacteriophage virions, pp. 347–357. *In* A. J. Dalton and F. Hagunay (eds.) Ultrastructure of Animal Viruses and Bacteriophages. Academic Press, New York.
3. Artymiuk, P. J., T. A. Ceska, D. Suck, and J. R. Sayers. 1997. Prokaryotic 5′–3′ exonucleases share a common core structure with gamma-delta resolvase. Nucleic Acids Res. 25:4224–4229.
4. Beckman, L. D., M. S. Hoffman, and D. J. McCorquodale. 1971. Pre-early proteins of bacteriophage T5: structure and function. J. Mol. Biol. 62:551–564.
5. Berget, S. M., T. J. Mozer, and H. R. Warner. 1976. Early events after infection of *Escherichia coli* by bacteriophage T5. II. Control of the bacteriophage-induced 5′-nucleotidase activity. J. Virol. 18:71–79.
6. Bhagwat, M. and N. G. Nossal. 2001. Bacteriophage T4 RNase H removes both RNA primers and adjacent DNA from the 5′ end of lagging strand fragments. J. Biol. Chem. 276:28516–28524.
7. Bonhivers, M., A. Ghazi, P. Boulanger, and L. Letellier. 1996. FhuA, a transporter of the *Escherichia coli* outer membrane, is converted into a channel upon binding of bacteriophage T5. EMBO J. 15:1850–1856.
8. Bonhivers, M. and L. Letellier. 1995. Calcium controls phage T5 infection at the level of the *Escherichia coli* cytoplasmic membrane. FEBS Lett. 374:169–173.
9. Bourguignon, G. J., T. K. Sweeney, and H. Delius. 1976. Multiple origins and circular structures in replicating T5 bacteriophage DNA. J. Virol. 18:245–259.
10. Bradbeer, C., M. L. Woodrow, and L. I. Khalifah. 1976. Transport of vitamin B_{12} in *Escherichia coli*: common receptor system for vitamin B_{12} and bacteriophage BF23 on the outer membrane of the cell envelope. J. Bacteriol. 125:1032–1039.
11. Bradley, D. E. 1963. The structure of coliphages. J. Gen. Microbiol. 23:435–445.
12. Bradley, D. E. 1967. Ultrastructure of bacteriophages and bacteriocins. Bacteriol. Rev. 31:230–314.
13. Ceska, T. A., J. R. Sayers, G. Stier, and D. Suck. 1996. A helical arch allowing single-stranded DNA to thread through T5 5′-exonuclease. Nature 382:90–93.
14. Chatterjee, D. K., R. K. Fujimura, J. H. Campbell, and G. F. Gerard. 1991. Cloning and overexpression of the gene encoding bacteriophage T5 DNA polymerase. Gene 97:13–19.
15. Chinnadurai, G. and D. J. McCorquodale. 1973. Requirement of a phage-induced 5′-exonuclease for the expression of late genes of bacteriophage T5. Proc. Natl. Acad. Sci. USA 70:3502–3505.
16. Chinnadurai, G. and D. J. McCorquodale. 1974. Regulation of expression of late genes of bacteriophage T5. J. Virol. 13:85–93.
17. Constantinou, A., K. Voelkel-Meiman, R. Sternglanz, M. M. McCorquodale, and D. J. McCorquodale. 1986. Involvement of host DNA gyrase in growth of bacteriophage T5. J. Virol. 57:875–882.
18. Decker, K., V. Krauel, A. Meesmann, and K. J. Heller. 1994. Lytic conversion of *Escherichia coli* by bacteriophage T5: blocking of the FhuA receptor protein by a lipoprotein expressed early during infection. Mol. Microbiol. 12:321–332.
19. Dervan, J. J., M. Feng, D. Patel, J. A. Grasby, P. J. Artymiuk, T. A. Ceska, and J. R. Sayers. 2002. Interactions of mutant and wild-type flap endonucleases with oligonucleotide substrates suggest an alternative model of DNA binding. Proc. Natl. Acad. Sci. USA 99:8542–8547.
20. Everett, R. D. 1981. DNA replication of bacteriophage T5. 3. Studies on the structure of concatemeric T5 DNA. J. Gen. Virol. 52:25–38.
21. Ferguson, A. D., E. Hofmann, J. W. Coulton, K. Diederichs, and W. Welte. 1998. Siderophore-mediated iron transport: crystal structure of FhuA with bound lipopolysaccharide. Science 282:2215–2220.
22. Ficht, T. A. and R. W. Moyer. 1980. Isolation and characterization of a putative bacteriophage T5 transcription-replication enzyme complex from infected *Escherichia coli*. J. Biol. Chem. 255:7040–7048.
23. Fox, J. W., A. Barish, C. E. Snyder, and R. Benzinger. 1982. Amino terminal sequence of the bacteriophage T5-coded gene A2 protein. Biochem. Biophys. Res. Commun. 106:265–269.
24. Frenkel, G. D. and C. C. Richardson. 1971. The deoxyribonuclease induced after infection of *Escherichia coli* by bacteriophage T5. I. Characterization of the enzyme as a 5′-exonuclease. J. Biol. Chem. 246:4839–4847.
25. Frenkel, G. D. and C. C. Richardson. 1971. The deoxyribonuclease induced after infection of *Escherichia coli* by bacteriophage T5. II. Role of the enzyme in replication of the phage deoxyribonucleic acid. J. Biol. Chem. 246:4848–4852.
26. Garforth, S. J., T. A. Ceska, D. Suck, and J. R. Sayers. 1999. Mutagenesis of conserved lysine residues in bacteriophage T5 5′–3′ exonuclease suggests separate mechanisms of endo and exonucleolytic cleavage. Proc. Natl. Acad. Sci. USA 96:38–43.
27. Gentz, R. and H. Bujard. 1985. Promoters recognized by *Escherichia coli* RNA polymerase selected by function: highly efficient promoters from bacteriophage T5. J. Bacteriol. 164:70–77.
28. Gentz, R., A. Langner, A. C. Chang, S. N. Cohen, and H. Bujard. 1981. Cloning and analysis of strong promoters is made possible by the downstream placement of a RNA termination signal. Proc. Natl. Acad. Sci. USA 78:4936–4940.
29. Haggard-Ljungquist, E., C. Halling, and R. Calendar. 1992. DNA sequences of the tail fiber genes of bacteriophage

P2: evidence for horizontal transfer of tail fiber genes among unrelated bacteriophages. J. Bacteriol. 174:1462–1477.

30. Heller, K. and V. Braun. 1979. Accelerated adsorption of bacteriophage T5 to *Escherichia coli* F, resulting from reversible tail fiber-lipopolysaccharide binding. J. Bacteriol. 139:32–38.
31. Heller, K. and V. Braun. 1982. Polymannose O-antigens of *Escherichia coli*, the binding sites for the reversible adsorption of bacteriophage T5+ via the L-shaped tail fibers. J. Virol. 41:222–227.
32. Heller, K. J. 1984. Identification of the phage gene for host receptor specificity by analyzing hybrid phages of T5 and BF23. Virology 139:11–21.
33. Heller, K. J. and V. Krauel. 1986. Cloning and expression of the *ltf* gene of bacteriophage T5. J. Bacteriol. 167:1071–1073.
34. Heller, K. J. and H. Schwarz. 1985. Irreversible binding to the receptor of bacteriophages T5 and BF23 does not occur with the tip of the tail. J. Bacteriol. 162:621–625.
35. Herman, R. C. and R. W. Moyer. 1974. In vivo repair of the single-strand interruptions contained in bacteriophage T5 DNA. Proc. Natl. Acad. Sci. USA 71:680–684.
36. Herman, R. C. and R. W. Moyer. 1975. In vivo repair of bacteriophage T5 DNA: an assay for viral growth control. Virology 66:393–407.
37. Heusterspreute, M., V. Ha-Thi, S. Tournis-Gamble, and J. Davison. 1987. The first-step transfer-DNA injection-stop signal of bacteriophage T5. Gene 52:155–164.
38. Hollingsworth, H. C. and N. G. Nossal. 1991. Bacteriophage T4 encodes an RNase H which removes RNA primers made by the T4 DNA replication system in vitro. J. Biol. Chem. 266:1888–1897.
39. Kaliman, A. V., A. I. Krutilina, and V. M. Kriukov. 1987. The structure of 2 promoters and a transcription terminator from the region of early genes *D10–D15* of bacteriophage T5. Mol. Gen. Mikrobiol. Virusol. 10:14–19.
40. Kaliman, A. V., A. I. Krutilina, V. M. Kryukov, and A. A. Bayev. 1986. Cloning and DNA sequence of the 5′-exonuclease gene of bacteriophage T5. FEBS Lett. 195:61–64.
41. Kaliman, A. V., V. M. Kryukov, and A. A. Bayev. 1988. The nucleotide sequence of the region of bacteriophage T5 early genes *D10–D15*. Nucleic Acids Res. 16:10353–10354.
42. Kaliman, A. V., V. E. Kulshin, M. G. Shlyapnikov, V. N. Ksenzenko, and V. M. Kryukov. 1995. The nucleotide sequence of the bacteriophage T5 *ltf* gene. FEBS Lett. 366:46–48.
43. Killmann, H., G. Videnov, G. Jung, H. Schwarz, and V. Braun. 1995. Identification of receptor binding sites by competitive peptide mapping: phages T1, T5, and phi 80 and colicin M bind to the gating loop of FhuA. J. Bacteriol. 177:694–698.
44. Krauel, V. and K. J. Heller. 1991. Cloning, sequencing, and recombinational analysis with bacteriophage BF23 of the bacteriophage T5 *oad* gene encoding the receptor-binding protein. J. Bacteriol. 173:1287–1297.
45. Labedan, B. and J. Legault-Demare. 1973. Penetration into host cells of naked, partially injected (post-FST) DNA of bacteriophage T5. J. Virol. 12:226–229.
46. Lambert, O., L. Plancon, J. L. Rigaud, and L. Letellier. 1998. Protein-mediated DNA transfer into liposomes. Mol. Microbiol. 30:761–765.
47. Lanni, Y. 1969. Functions of two genes in the first-step-transfer DNA of bacteriophage T5. J. Mol. Biol. 44:173–183.
48. Lanni, Y. T. 1968. First-step-transfer deoxyribonucleic acid of bacteriophage T5. Bacteriol. Rev. 32:227–242.
49. Lanni, Y. T. and D. J. McCorquodale. 1963. DNA metabolism in T5-infected *Escherichia coli*: biochemical function of a presumptive genetic fragment of the phage. Virology 19:72.
50. Leavitt, M. C. and J. Ito. 1989. T5 DNA polymerase: structural–functional relationships to other DNA polymerases. Proc. Natl. Acad. Sci. USA 86:4465–4469.
51. Letellier, L., K. P. Locher, L. Plancon, and J. P. Rosenbusch. 1997. Modeling ligand-gated receptor activity. FhuA-mediated ferrichrome efflux from lipid vesicles triggered by phage T5. J. Biol. Chem. 272:1448–1451.
52. McCorquodale, D. J. 1975. The T-odd bacteriophages. CRC Crit. Rev. Microbiol. 4:101–159.
53. McCorquodale, D. J., C. W. Chen, M. K. Joseph, and R. Woychik. 1981. Modification of RNA polymerase from *Escherichia coli* by pre-early gene products of bacteriophage T5. J. Virol. 40:958–962.
54. McCorquodale, D. J., J. Gossling, R. Benzinger, R. Chesney, L. Lawhorne, and R. W. Moyer. 1979. Gene D5 product of bacteriophage T5: DNA-binding protein affecting DNA replication and late gene expression. J. Virol. 29:322–327.
55. McCorquodale, D. J., A. R. Shaw, P. K. Shaw, and G. Chinnadurai. 1977. Pre-early polypeptides of bacteriophages T5 and BF23. J. Virol. 22:480–488.
56. McCorquodale, D. J. and H. Warner. 1988. Bacteriophage T5 and related phages, pp. 439–475. *In* R. Calendar (ed.) The Bacteriophages. Plenum Press, New York.
57. Mikoulinskaia, G. V., S. I. Gubanov, A. Zimin, I. V. Kolesnikov, S. A. Feofanov, and A. I. Miroshnikov. 2003. Purification and characterization of the deoxynucleoside monophosphate kinase of bacteriophage T5. Protein Expr. Purif. 27:195–201.
58. Mondigler, M., T. Holz, and K. J. Heller. 1996. Identification of the receptor-binding regions of pb5 proteins of bacteriophages T5 and BF23. Virology 219:19–28.
59. Mondigler, M., R. T. Vogele, and K. J. Heller. 1995. Overproduced and purified receptor binding protein pb5 of bacteriophage T5 binds to the T5 receptor protein FhuA. FEMS Microbiol. Lett. 130:293–300.
60. Moyer, R. W. and C. T. Rothe. 1977. Role of the T5 gene *D15* nuclease in the generation of nicked bacteriophage T5 DNA. J. Virol. 24:177–193.
61. Mozer, T. J., R. B. Thompson, S. M. Berget, and H. R. Warner. 1977. Isolation and characterization of a bacteriophage T5 mutant deficient in deoxynucleoside 5′-monophosphatase activity. J. Virol. 24:642–650.
62. Nakayama, S., T. Kaneko, H. Ishimaru, H. Moriwaki, and K. Mizobuchi. 1994. Cloning, sequencing, and expression of bacteriophage BF23 late genes *24* and *25* encoding tail proteins. J. Bacteriol. 176:7280–7290.

63. Nichols, B. P. and J. E. Donelson. 1977. Sequence analysis of the nicks and termini of bacteriophage T5 DNA. J. Virol. 22:520–526.
64. Pedruzzi, I., J. P. Rosenbusch, and K. P. Locher. 1998. Inactivation in vitro of the *Escherichia coli* outer membrane protein FhuA by a phage T5-encoded lipoprotein. FEMS Microbiol. Lett. 168:119–125.
65. Rhoades, M. 1982. New physical map of bacteriophage T5 DNA. J. Virol. 43:566–573.
66. Rhoades, M. and E. A. Rhoades. 1972. Terminal repetition in the DNA of bacteriophage T5. J. Mol. Biol. 69:187–200.
67. Rhoades, M., J. Schwartz, and J. M. Wahl. 1980. New deletion mutant of bacteriophage T5. J. Virol. 36:622–626.
68. Rice, A. C., T. A. Ficht, L. A. Holladay, and R. W. Moyer. 1979. The purification and properties of a double-stranded DNA-binding protein encoded by the gene D5 of bacteriophage T5. J. Biol. Chem. 254:8042–8051.
69. Rogers, S. G. and M. Rhoades. 1976. Bacteriophage T5-induced endonucleases that introduce site-specific single-chain interruptions in duplex DNA. Proc. Natl. Acad. Sci. USA 73:1576–1580.
70. Rose, J. D. and D. J. McCorquodale. 1990. Identity of genes *A2* and *A3* of bacteriophage BF23. Virology 177:753–755.
71. Saigo, K. 1978. Isolation of high-density mutants and identification of nonessential structural proteins in bacteriophage T5; dispensability of L-shaped tail fibers and a secondary major head protein. Virology 85:422–433.
72. Sayers, J. R. and F. Eckstein. 1990. Properties of overexpressed phage T5 D15 exonuclease. Similarities with *Escherichia coli* DNA polymerase I 5′–3′ exonuclease. J. Biol. Chem. 265:18311–18317.
73. Sayers, J. R. and F. Eckstein. 1991. A single-strand specific endonuclease activity copurifies with overexpressed T5 D15 exonuclease. Nucleic Acids Res. 19:4127–4132.
74. Shaw, A. R. and J. Davison. 1979. Polarized injection of the bacteriophage T5 chromosome. J. Virol. 30:933–935.
75. Shaw, A. R., D. Lang, and D. J. McCorquodale. 1979. Terminally redundant deletion mutants of bacteriophage BF23. J. Virol. 29:220–231.
76. Shlyapnikov, M. G., V. N. Ksenzenko, V. M. Kryukov, and A. A. Bayev. 1995. Specific properties of phage T5-encoded tRNAs. Mol. Biol. 28:818–823.
77. Smith, M. C., N. Burns, J. R. Sayers, J. A. Sorrell, S. R. Casjens, and R. W. Hendrix. 1998. Bacteriophage collagen. Science 279:1834.
78. Snyder, C. E., Jr. 1984. Bacteriophage T5 gene A2 protein alters the outer membrane of *Escherichia coli*. J. Bacteriol. 160:1191–1195.
79. Snyder, C. E., Jr. 1991. Amino acid sequence of the bacteriophage T5 gene A2 protein. Biochem. Biophys. Res. Commun. 177:1240–1246.
80. Snyder, C. E., Jr. and R. H. Benzinger. 1981. Second-step transfer of bacteriophage T5 DNA: purification and characterization of the T5 gene A2 protein. J. Virol. 40:248–257.
81. Szabo, C., B. Dharmgrongartama, and R. W. Moyer. 1975. The regulation of transcription in bacteriophage T5-infected *Escherichia coli*. Biochemistry 14:989–997.
82. Szabo, C. and R. W. Moyer. 1975. Purification and properties of a bacteriophage T5-modified form of *Escherichia coli* RNA polymerase. J. Virol. 15:1042–1046.
83. von Gabain, A. and H. Bujard. 1977. Interaction of *E. coli* RNA polymerase with promotors of coliphage T5: the rates of complex formation and decay and their correlation with in vitro and in vivo transcriptional activity. Mol. Gen. Genet. 157:301–311.
84. von Gabain, A. and H. Bujard. 1979. Interaction of *Escherichia coli* RNA polymerase with promoters of several coliphage and plasmid DNAs. Proc. Natl. Acad. Sci. USA 76:189–193.
85. Warner, H. R., R. F. Drong, and S. M. Berget. 1975. Early events after infection of *Escherichia coli* by bacteriophage T5. Induction of a 5′-nucleotidase activity and excretion of free bases. J. Virol. 15:273–280.
86. Wiest, J. S. and D. J. McCorquodale. 1990. Characterization of pre-early genes in the terminal repetition of bacteriophage BF23 DNA by nucleotide sequencing and restriction mapping. Virology 177:745–754.
87. Zweig, M. and D. J. Cummings. 1973. Structural proteins of bacteriophage T5. Virology 51:443–453.
88. Zweig, M., H. S. Rosenkranz, and C. Morgan. 1972. Development of coliphage T5: ultrastructural and biochemical studies. J. Virol. 9:526–543.

20

The T7 Group

IAN J. MOLINEUX

Bacteriophage T7 is the prototype member of the T7 group of phages that are traditionally thought to be obligately lytic. Although isolated in 1945, early studies on T3 and T7 were much less extensive than those on the T-even phages. However, in 1969 a comprehensive genetic map of T7 was published (269, 276), several phage-coded enzymes involved in DNA replication and transcription were purified and studied biochemically (22, 27, 85, 96), and T7 rapidly became a focus of attention. These studies have continued to the present, the mechanism of transcription by T7 RNA polymerase is known in exquisite detail, and in vitro DNA replication catalyzed by T7 proteins is one of the best understood model systems. In more recent times, T7 RNA polymerase-based expression systems (279; also see chapter 43), DNA sequencing using Sequenase, and a T7 display system (229), have all made people extremely familiar with T7 parts, if not with the phage itself. The ease of laboratory manipulation of T7, its fast growth rate, and of course the wealth of information that has accrued in the past decades, has also made T7 a prime model system for experimental molecular evolution and evolutionary genetics (17–19, 48, 100).

Overview

Bacteriophage T7, together with its relative T3, was originally isolated as a member of the seven Type phages that grow on *E. coli* B (53). Plaques of T7 and T3 (but not more distantly related phages) are characteristically large on most *E. coli* strains, and they continue to expand on extended incubation (319). Unlike T1, which also makes large plaques, T7 and T3 do not cause persistent contamination in the laboratory as they die rapidly on drying. On wild-type hosts T7 plaque size generally decreases with decreasing incubation temperature, the relative efficiency of plating (eop) remains constant from 43°C to about 15°C, and an uncharacterized mutant forms plaques at 10°C at an eop ~1 (P. Kemp and I. J. Molineux, unpublished observations).

Several T7-like coliphages have been isolated that by morphology and hybridization studies are very closely related to T7 itself (e.g., figure 20-1A). Genome sequences of the closeknit T7 group are thus far limited to T7 and T3, the yersiniophages ϕA1122 and ϕYeO3-12, and the *Pseudomonas putida* phage gh-1 (62, 78, 135, 206, 207). All five exhibit a highly conserved genome organization and, at the sequence level, differ mainly in the presence or absence of several nonessential genes, many of which exhibit homology to homing endonucleases or other parasitic elements. In accord with the modular evolution of phages, it is becoming apparent that more distant relatives of T7 are abundant whose genomes are organized differently and only a few of whose genes are clearly homologous to T7. Among these are the dual capsule-specific coliphage K1-5 and the *Salmonella* phage SP6 (56, 240), the *P. aeruginosa* phage ϕKMV (144), the *Xanthomonas campestris* pv. *oryzae* phage Xp10 (323), and the marine cyanophage P60 (66). The roseophage SIO1 (224) and the *Vibrio parahaemolyticus* phage VpV262 (90) are even further removed from T7. Interestingly, the *Agrobacterium tumefaciens* chromosome contains proteins with extensive similarity to T7 RNA polymerase (RNAP) and the T7 tail fiber but no other easily recognized T7-like genes (GenBank AE007869). Even more surprising is the *Pseudomonas putida* KT2440 chromosome, which contains a putative prophage, almost all of whose genes show greatest similarity to T7, T3, or ϕYeO3-12 (201; GenBank NC_002947). Most essential genes of T7 are present in the prophage, notable exceptions being homologs to genes *0.7* and *2*, which modify the host transcription and translational machinery, and the *17.5* holin. No repressor protein is obvious within the prophage and, understandably, host promoters that would allow expression of the phage RNAP are missing, as determined by sequence inspection. The prophage is flanked by a repeat sequence that inserted into tRNA-leu-1, and also contains a putative integrase. The presence of this almost complete copy of a T7-like genome as a cryptic prophage complete with a putative integrase raises many questions regarding the evolution and ecology of what are referred to as obligate lytic phages.

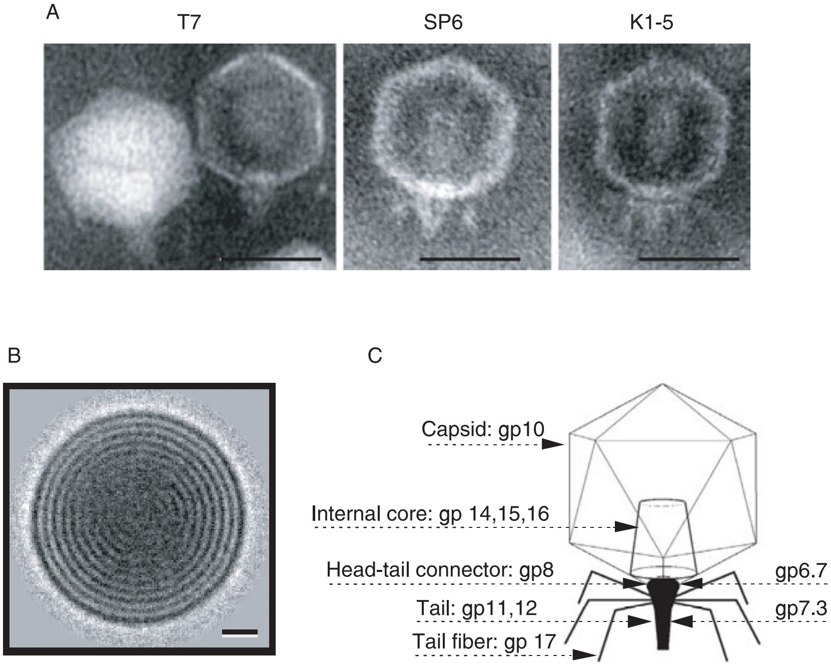


Figure 20-1 Phage T7 and T7-like virions. A: Electron micrographs of T7, SP6, and K1-5. The internal core is visible but in T7 and K1-5 it has detached from the head–tail connector. Note also the different tail structure between T7 and SP6 or K1-5. The thin T7 tail fibers are not apparent. Scale bars represent 50 nm. Courtesy of S. Adhya and D. Scholl. B: Cryo-electron micrograph of a tailless T7 head showing the "concentric ring" pattern of the densely packed spooled DNA. The scale bar represents 10 nm. Courtesy of M. Cerritelli. C: Schematic of the T7 virion indicating its protein components.

As more sequence information becomes available, any distinct differences between T7 or its immediate family and more distantly related phages will likely become even more blurred. However, to date no T7-like phage has been found to infect Gram-positive bacteria. This chapter will focus mainly on T7 although significant differences that are known in other phages will be emphasized. Earlier reviews, especially those by Krüger and Schroeder (137) and Dunn and Studier (62), should also be consulted.

The primary distinguishing features of the T7 group have been that they are members of the *Podoviridae* and code for a rifampicin-resistant, single-polypeptide RNAP that is responsible for most phage gene expression. However, this classification may need re-evaluation as Xp10 contains a T7-related RNAP but has λ morphology, and the T7-like marine phages SIO1 and VpV262 lack a phage-specific RNAP. In addition to host range differences, the closer relatives of T7 can be grouped by the promoter specificity of the phage polymerase. Members of the same group have a common promoter sequence; these phages recombine extensively with each other but much less efficiently with members of a different group.

Aside from the translational apparatus and the biosynthetic machinery for precursor synthesis, T7 growth is remarkably independent of host enzymes. *E. coli* RNAP is used to make early RNAs, but most transcription is catalyzed by the T7 enzyme and, except for thioredoxin, both DNA replication and recombination are independent of host proteins.

The T7 Virion

A spherical approximation of the icosahedron, $T = 7$, T7 capsid has a diameter of 60–61 nm (figure 20-1A), enclosing a volume of about 10^5 nm^3 (227, 267). The T7 genome occupies about 45% of the head volume. The 40 kb double-stranded DNA is in B-form, has a center-to-center helix spacing of 2.4 nm, and is spooled around the internal core in about six coaxial shells (23) (figure 20-1B). A genome as large as 103% and as small as 85% unit length can be packaged into infectious particles. The capsid shell consists of 415 molecules of gp10, about 95% of which is normally the major capsid protein, gp10A. The remainder is the minor capsid protein, gp10B, which results from a −1 reading-frame shift near the 3′ end of gene *10A* (44, 45, 62). The growth of mutant phages that make only gp10A is indistinguishable from wild-type, but plaques of phages that make only gp10B are smaller than usual. T3 (43, 44, 45), and several other members of the T7 group, are also thought to produce a translationally shifted second capsid protein (95), although phages gh-1, SP6, and K1-5 do not (135, 240).

At one vertex of the gp10 icosahedron the dodecameric head–tail connector (gp8) is inserted (figure 20-1C).

Expression of gene *8* clones results in mixtures of oligomers that contain 12 or 13 molecules of gp8 (25, 132, 298), but only 12mers are found in mature virions or within filled heads. The cryo-electron microscopic structure of the dodecameric cloned T3 connector reveals a similar structure to the ϕ29 and SPP1 connectors, a 12-lobed wide domain that is inserted into the prohead and a narrower domain that interacts with the tail (297, 298). An open channel runs through the cloned connector, but the channel is closed when the connector is isolated from virions (57). In the latter work the channel through which the genome passes has an average diameter of about 3.7 nm but with a constriction down to about 2.2 nm. The constriction is close to that of the diameter of duplex DNA, and may be one reason why, in contrast to many other phages, there is no transient leakage of ions following T7 infection (138). The change in diameter of the channel may be due to conformational changes in the connector associated with DNA packaging during assembly or (perhaps in addition) to incorporation of gp6.7 into the head. The product of gene *13* was once thought to be part of the virion but is actually required to incorporate gp6.7 (which migrates similarly to gp13 in many SDS gel systems) into heads. Gp6.7 may be essential, but morphologically normal particles are formed in its absence.

Three other head proteins form an internal, cylindrical structure about 26 nm long by 21 nm wide that is attached to, and coaxial with, the 12-fold symmetrical connector (241, 243). The core consists of coaxially stacked rings and exhibits 8-fold symmetry (26); 4 copies of gp16, 8 of gp15, and 8–12 of gp14 are currently thought to comprise the structure. The head also contains 15–20 copies of gp6.7 and a small amount of the nonessential gp7 (P. Kemp and I. J. Molineux, unpublished observations). Their location with respect to the core is not yet known (figure 20-1C).

The stubby T7 tail is 23 nm long, tapering from 21 nm wide at the connector to 9 nm at the distal end (262). The tail exhibits 6-fold symmetry and is estimated to consist of 12 copies of gp11 and 6 copies of gp12. About 30 copies of gp7.3 are also part of the tail but their precise location has not been determined. The six tail fibers, each consisting of three parallel gp17 molecules, are attached, through an N-terminal domain of the protein, near the top of the tail (115, 263).

Genetic Organization

The T7 genome contains 39,937 bp and codes for 56 known or potential genes (62, 185) (table 20-1; figure 20-2). A 160 bp direct repeat lies at each end of the T7 genome, the length and sequence of the terminal repeat (TR) varying among different members of the T7 group. Although circular molecules can be formed in vitro after exonuclease digestion, they are not found in vivo. Rather, linear concatemers, containing only a single copy of TR between genomes, are the initial products of DNA replication.

T7 DNA contains only the four normal bases. A limited amount of adenine methylation occurs from the type I restriction-modification enzyme and from Dam and Dcm methylases, but not all potential sites are fully methylated because the rate of T7 DNA replication and packaging is so high. Some of the other T7-like phage DNAs, including that of T3, are completely unmethylated since they code for an *S*-adenosylmethionine hydrolase, which removes the methyl group donor.

Sequences corresponding to many type II restriction enzymes with palindromic recognition sites are grossly underrepresented in the genomes of the T7 phage family. Avoidance of palindromic sequences is the primary mechanism used to avoid restriction by type II enzymes. T7 is also sensitive to restriction by type III enzymes and thus does not grow in P1 R^+M^+ lysogens. However, T7 growth is insensitive to *Eco*P15 as the restriction enzyme requires two copies of 5′-CAGCAG/5′-CTGCTG in opposite orientations (176), and all 36 copies of this hexamer have the same polarity in the T7 genome. T3 is restricted by *Eco*P15 since its genome contains the recognition sequence in both orientations. The genomes of all immediate members of the T7 group examined to date contain remnants of homing endonucleases or related mobile parasitic elements. Although some are inserted within coding regions, most are present as apparently precise insertions between genes but they are not always present at the same locations in different genomes. These genetic elements have no known function in phage development, although a frameshift mutation in T7 gene *2.8* is translationally polar on gene *3* (225).

In vitro, the *Escherichia coli* RecBCD nuclease rapidly degrades the linear T7 genome. T7 gp5.9 binds stoichiometrically to RecBCD holoenzyme, inhibiting both its nuclease and ATPase activities (153). However, gene *5.9* is nonessential in RecBCD-containing cells, and gp5.9 is not found in the phage particle. The infecting T7 genome is most likely protected from RecBCD by one or more of the proteins ejected from the virion or by a cellular component.

All transcription is from left to right on the T7 genetic map; genes and regulatory elements are defined by their position on the nontranscribed strand (figure 20-2). Genes are numbered sequentially in the order that they are transcribed initially; T7 promoters (prefix ϕ), and RNase III processing sites (prefix R) are usually named by the gene immediately downstream. Whole-number genes were originally defined by complementation tests as being essential (269), whereas most noninteger genes were first defined as being nonessential or as having conditionally essential activities. More recent studies have shown that genes *2.5*, *6.7*, and *7.3* may be considered essential, and that gene *7* is nonessential (table 20.1). In addition, but for unknown reasons, gene *2* is not essential on *E. coli* C strains (273).

Table 20.1 T7 genes

Class	Function[a]	Selected references
Class I		
0.3	B-DNA mimic; anti-type I restriction	301
0.4, 0.5, 0.6A, 0.6B	Not conserved; nonessential	
0.7	Protein kinase; host-transcription shutoff; Col Ib exclusion 84, 211	84
211		
1	T7 RNA polymerase	159, 287, 321, 329
1.1	Conserved, nonessential	
1.2	*E. coli* dGTPase inhibitor; F-exclusion	188, 190, 233, 235, 237
1.3	DNA ligase	271, 262
Class II		
1.4, 1.5, 1.6	Not conserved, nonessential	
1.7	Full-length gene not conserved; beneficial for growth	
1.8	Poorly conserved, nonessential	
2	*E. coli* RNAP inhibitor	200
2.5	SSB	128, 133
2.8	Not conserved, nonessential; homing endonuclease?	
3	Endonuclease I, Holliday junction resolvase	51, 122
3.5	Amidase (lysozyme); regulates T7 RNAP activity	329
3.8	Not conserved, nonessential; homing endonuclease	
4A	Primase-helicase; gp4B helicase from internal in-frame start	69
4.1, 4.2	Overlappons; not conserved	
4.3, 4.5	Conserved, nonessential	275
4.7	Not conserved, nonessential	275
5	DNA polymerase	58, 148, 149
5.3	Not conserved, nonessential, homing endonuclease	
5.5	Conserved, nonessential, binds *E. coli* HNS; λ*rex* exclusion Non-conserved −1 frameshift leads to T7 *5.5–5.7* fusion	153, 154
5.7	Conserved, nonessential	
5.9	Inhibits RecBCD nuclease, nonessential, not conserved	153
6	5′→3′double-stranded exonuclease, RNase H	147, 251
6.3	Poorly conserved, nonessential	
Class III		
6.5	Conserved, nonessential	
6.7	Virion protein; ejected into infected cell	119
7	Nonessential, not conserved; host range	62, 273
7.3	Essential virion protein; ejected into infected cell	(P. Kemp and I. J. Molineux, unpublished observations)
7.7	Not conserved; homing endonuclease	
8	Head–tail connector protein	25, 298
9	Scaffolding protein	24
10A	Major capsid protein; −1 frameshift yields minor capsid protein gp10B F exclusion	24, 45, 188, 190
11	Tail protein	262
12	Tail protein	262
13	Essential; required for gp6.7 incorporation in virion	(P. Kemp and I. J. Molineux, unpublished observations)
14	Internal core protein; ejected into infected cell	189
15	Internal core protein; ejected into infected cell	189
16	Internal core protein; ejected into infected cell	183, 184, 189, 268
17	Tail fiber protein	115, 263
17.5	Class II holin	302
18	Small terminase subunit	89, 309
18.5–18.7	Conserved; λ *Rz-Rz1* homologs	21
19	Large terminase subunit	193, 195, 310
19.2, 19.3	Overlappons, conserved	
19.5	Nonessential, conserved	126, 127

[a] Conserved or not conserved refers to close relatives of T7.

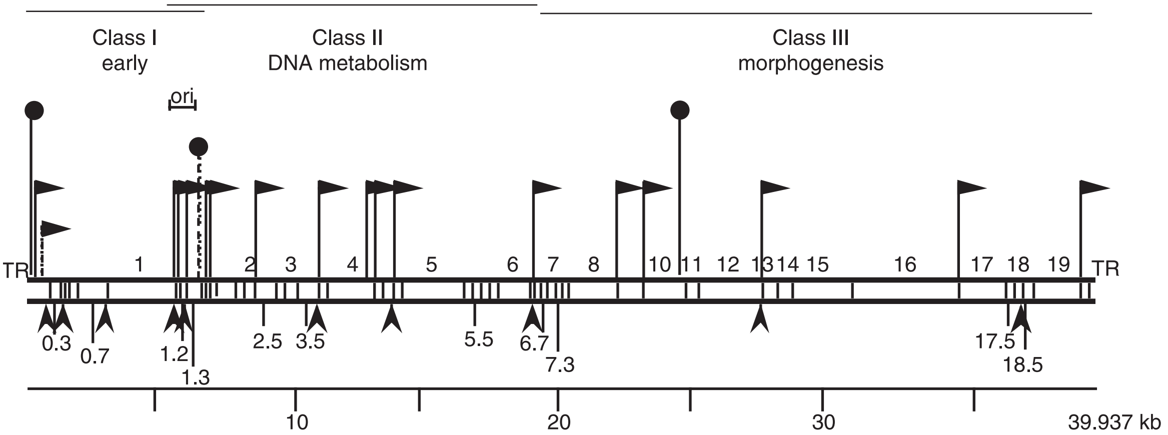


Figure 20-2 Schematic of the genetic organization of bacteriophage T7. Gene expression is from left to right and is temporally controlled, that is left-most genes are transcribed first, right-most genes last. Only genes with known function are named. T7 promoters are indicated by horizontal arrowheads on continous vertical lines. The three major *E. coli* promoters (combined into one) are indicated by an arrowhead on a dashed vertical line. Transcription terminators are represented by filled circles for *E. coli* RNA polymerase (dashed line) and T7 RNA polymerase (continous); the primary origin of DNA replication is indicated. Arrowheads below the map represent sites of RNase III processing. For the exact locations of all genetic elements see Dunn and Studier (62) and Genbank NC_001604.

Genes are close-packed, and about 92% of the genome is coding sequence; the arrangement of close-packed and slightly overlapping genes in T7 may make translational coupling quite common. There are five cases of substantial or complete gene overlap, where the second gene is in a different reading frame. Genes *4.1* and *4.2* lie almost entirely within gene *4*, gene *18.7* lies within gene *18.5*, and genes *19.2* and *19.3* both overlap gene *19*. In addition, gene *4* codes for two proteins via an internal in-frame initiation; the longer gp4A is both a primase and a helicase whereas gp4B has only helicase activity. Three cases of programmed translational frameshifting have been proposed in T7, but only the −1 shift that produces the gp10B protein has been characterized (44, 45, 62).

A time course of protein synthesis reveals three classes of gene products; class I, or early, genes are synthesized from about 2 to 8 minutes post-infection, class II genes from about 6 to 15 minutes, and class III genes from 8 minutes until lysis, which begins after 25 minutes at 30°C (270). The 10 class I genes, *0.3–1.3*, are transcribed by *E. coli* RNAP but only one, T7 RNAP, is essential. Genes *1.1* through *6.3* comprise the class II region and genes *6.5–19.5* the class III region. Both class II and III genes are transcribed by T7 RNAP. In general, class I genes serve to establish favorable conditions in the cell for phage development, class II genes are involved in DNA metabolism, and class III genes have DNA packaging, virion assembly, and lysis functions.

Adsorption and Host Range

The host range of wild-type T7 appears to be restricted to *E. coli*, rough *S. typhimurium* LT2, and various *Shigella* strains (T7 undergoes an abortive infection in *Shigella sonnei* D2-3748; 6, 94, 212). Extended host range mutants of T7 also allow growth in *Yersinia pestis* and *Yersinia pseudotuberculosis*, and vice versa. Comparable mutants of *Y. pestis* ϕA1122 grow in *E. coli* (78; E. Ramunculov, M. C. Chu and I. J. Molineux, unpublished observations). Extended or altered host range mutants of the T7 group of phages are readily isolated, and "adaptation" of phages to a variety of different hosts was once a common practice (e.g., 145). This practice, whether intentional or otherwise, has resulted in some confusion. The *Y. pestis* phage ϕA1122 has been described as a coliphage without indication that a host range mutant was being used, and even T7 and T3 have been misidentified by more than one laboratory (274). As many spontaneous extended host range mutants infect their new host less efficiently than they infect their original host, any inadvertent use of mutants can be misleading. Tail fiber interaction with lipopolysaccharide (LPS) is a necessary early step in infection, and as LPS structures are themselves rather variable, the tail fiber is likely a malleable structure that easily adapts to a variety of receptors.

On *E. coli* B strains, T3 adsorbs to the penultimate glucose of the LPS (213), the same residue as T4. As a consequence, B/3 strains, which have lost both terminal glucose molecules, are also B/4 but remain T7^{S}; the initial receptor for T7 is deeper within the inner LPS core and B/7 strains are obligatorily B/3,4. T7 grows normally on most *E. coli* C and K-12 derivatives, T3 also grows on *E. coli* C but fails to adsorb to many common laboratory K-12 strains. Extended host range mutants with an altered tail fiber (gp17) overcome the adsorption defect. Neither T7 nor T3 form plaques on smooth *E. coli* strains that contain a complete outer core LPS, nor do they plaque on capsulated strains. However, the

Salmonella phage SP6 grows on both rough and smooth strains. Phages related to T7 or SP6 grow on various K antigen-containing *E. coli* strains (39, 140, 141, 155, 210, 239). Remarkably, each virion of the SP6-related coliphage ϕK1-5 contains two sets of tail fibers, one of which adsorbs to the K1, the other to the K5 capsule (239). The amino acid sequence of the tail fiber of ϕK1F is related to that of T7 at its N-terminal end, the middle domain, which is homologous to the tail fibers of other K1-specific phages, contains the endosialidase motifs. The sequence of the *Klebsiella* phage K11 tail fibers is not yet known but it has been shown that the capsule-degrading activity resides in the tail structure (231). The C-terminal sequences of the ϕK1F tail fiber exhibit patches of similarity to other tail fibers, including those of morphologically unrelated phages, an example of mosaicism among phage adsorption organelles (87).

The conservation of N-terminal tail fiber sequences among T7 group members is consistent with this region of the fiber binding to the tail (115, 263), and sequence divergence at the C-terminal end suggests that the latter corresponds to the LPS-binding ligand. This simple idea fits with the mutational sites in T3 host range mutants (selected on *E. coli* strains that do not adsorb $T3^+$) being located close to the 3′ end of the gene (M. I. Pajunen, I. J. Molineux and M. Skurnik, unpublished observations). However, some ϕA1122 extended host range mutants that grow on both *E. coli* and *Y. pestis* are altered near the N-terminus of gp17 (78).

Distinct from the failure to adsorb or the susceptibility to endogenous restriction enzymes, T7 fails to productively infect a number of natural *E. coli* strains. In some cases, normally nonessential genes allow growth and thus only the mutant phage is excluded, but growth in *S. sonnei* D2-3748 requires T7 to contain a missense mutation in the gene for the major capsid protein (212). The best understood T7 abortive infection, which bears some similarities to the classic system of T4 *r*II mutant exclusion by λ rex^+ lysogens, is the failure of T7 to grow in cells containing the conjugative plasmid F (164). F exclusion is found using most close relatives of T7 (including SP6; 331) but, notably, not T3.

DNA Penetration

After tail fiber attachment to LPS, the tip of the tail likely binds a cellular component, but any specificity to the interaction has not been demonstrated. However, an expansion of the phage SP6 host range is conferred by mutant tail proteins (J. Kieleczawa and I. J. Molineux, unpublished observations), suggesting that an interaction may be an important step in the adsorption process. Gp7.3 and gp6.7 are ejected from the virion into the outer membrane where they get degraded (120). This step likely corresponds to the beginning of the eclipse phase as particles lacking either protein are noninfectious. The internal core structure then disaggregates. Its components pass through the connector and the tail and then enter the cell (189). The gp14 molecules ejected from the infecting particle have been localized to the outer membrane, where they presumably form a channel, but no structural information is currently available. Gp16 is likely the next protein to enter the cell, with its N-terminal sequences leading. The N-terminal portion of gp16 is homologous to the catalytically active region of *E. coli* soluble lytic transglycosylase (183) and gp16 has been shown to possess peptidoglycan hydrolase activity (184). This activity is essential for phage growth when cells are in the late logarithmic stage of growth or are growing at low temperature. The natural crosslinking of the Gram-negative cell wall is a barrier to diffusion of molecules greater than approximately 50 kDa (52). The conditions where gp16 hydrolase activity is essential for phage growth are those where the peptidoglycan is thought to be more highly crosslinked and efficient infection requires enlarging a hole across the cell wall.

Membrane fractionation of T7-infected cells suggests that the gp15 and gp16 ejected from the infecting particle span both the outer and cytoplasmic membranes. It is likely that the N-terminal region of gp16 abuts the presumptive gp14 channel across the outer membrane and extends across the periplasm (P. Kemp, C.-Y. Chang, L. R. Garcia and I. J. Molineux, unpublished observations). Although topological studies have not been conducted, it is currently thought that a C-terminal domain of gp16 spontaneously inserts into the cytoplasmic membrane from its outer face, thereby making a channel for DNA translocation. The C-terminus of gp16 interacts with gp15 (C.-Y. Chang and I. J. Molineux, unpublished observations), which suggests that the latter may be proximal to the cell cytoplasm. Both gp15 and gp16 are large proteins (84.2 and 143.8 kDa, respectively) and therefore cannot maintain an extensive tertiary structure as they exit the virion, through the connector and tail, in order to enter the cell. About 50% of gp16 is predicted to be made up of short α-helices interspersed by random coils. It was suggested that gp16 might be comparable to a folding carpenter's rule, where each segment of the rule corresponds to an α-helix. By opening out from a folded rule conformation, gp16 may be able to exit the capsid as an extended molecule while retaining sufficient structural information to refold rapidly in the periplasm (189).

The combination of gp14, gp15, and gp16 thus forms a channel from the tip of the phage tail across the entire cell envelope, providing the genome access to the cell cytoplasm. T7 and its close relatives, which all code for homologous proteins, may therefore be described as having extensible tails, in contrast to the contractile tails of the *Myoviridae*.

When the trans-envelope channel has formed, about 850 bp of the 40 kb T7 genome enters the cell. There is a block to further efficient genome internalization that is normally removed by transcription (79, 81). The value of 850 bp results from the plating behavior, at temperatures below 30°C, of T7 deletion mutants that lack the major promoters (A1, A2, A3), and is comparable to estimates obtained in earlier studies at 30°C and 37°C using other approaches (187, 324, 325, 326). At 30°C, the block in the initial transcription-independent DNA translocation is not absolute, since at least 4 kb of all infecting genomes enter a rifampicin-treated cell within about 30 minutes (268), and after 24 hours of infection about 20% of infecting genomes are completely internalized. At least two additional major barriers to this transcription-independent DNA translocation are thought to exist. There are no known DNA sequences that create a barrier to genome internalization. The block after 850 bp has entered is thought to result from a measurement of DNA length (81), whereas the other major barriers to DNA translocation have not been precisely located. Even after 24 hours of infection only about 20% of infecting genomes are completely internalized in rifampicin-treated cells.

Mutations affecting a central region of gp16 allow complete genome entry in the absence of any transcription at a constant rate of approximately 70–75 bp per second at 30°C (268). The mutants were isolated by selection for a gain-of-function phenotype but all mutant proteins direct the same rate of genome internalization, suggesting that they have simply lost the ability to block transcription-independent DNA translocation. The rate of DNA translocation is highly temperature-dependent and the data fit Arrhenius kinetics—a result that is inconsistent with the process being driven by entropy or by energy stored in the packaged DNA within the virion (118). Particles containing an altered gp15 may exhibit modified kinetics of genome internalization, but it is not yet known exactly how (or whether) gp15 and gp16 function together in translocating DNA.

DNA translocation also requires a membrane potential. If the latter is collapsed while the genome is actively traversing the cytoplasmic membrane, then no further DNA translocation is possible. Thus, the membrane potential appears to provide the energy necessary to move DNA from the virion into the cell, but it is also necessary for formation and stability of the channel across the cytoplasmic membrane. Although neither gp15 nor gp16 contain ATP-binding motifs, it is not clear whether ATP is also necessary for transcription-independent DNA translocation from the virion across the membrane.

The leading 850 bp of T7 that enters the cell by the above process contains the T7 ϕOL promoter and the strong *E. coli* promoters, A1, A2, and A3. Transcripts from T7 ϕOL have not been detected during infection (T3 ϕOL transcripts are seen; 1) but ϕOL can direct genome entry if T7 RNAP is present in the cell prior to infection (79, 187). Normally, however, *E. coli* RNAP recognizes one or more promoters by transcribing the class I region and internalizes at least 7 kb of the genome at approximately 40 bp per second at 30°C. All three promoters are used at about the same efficiency in vivo under normal laboratory conditions (60, 272). In vitro, A3 is the strongest promoter at low temperature but at 37°C A1 is strongest (49). Perhaps A2 is most active under other, nonstandard conditions, and the three different promoters ensure efficient genome entry (as well as messenger RNA (mRNA) synthesis) under a variety of environmental conditions. Although *E. coli* RNAP is normally considered to terminate transcription at the early gene terminator, TE, termination is not 100% efficient and the entire T7 genome can be internalized using transcription by the host enzyme. A consensus *boxA* sequence lies downstream of the promoters, and *nus*-dependent transcription antitermination at *boxA* is necessary for complete genome internalization by RNAP in the absence of phage protein synthesis (W. P. Robins and I. J. Molineux, unpublished observations). Neither *nusA*, *nusB*, nor *boxA* is essential for normal phage growth, but an antitermination system that also reduces RNAP pausing could explain why the rate of T7 genome internalization by *E. coli* RNAP is faster than the average rate of cellular mRNA synthesis. Multiple transcripts originating from the A promoters allow synthesis of T7 RNAP, which normally completes the genome entry process at an estimated 200–300 bp per second at 30°C (79).

Complete internalization of the T7 genome in a productive infection takes about 10 minutes, approximately twice the time calculated from the measured rates of DNA translocation. However, this calculated value does not take into account the time taken to assemble the cell envelope channel for DNA transport nor any delay in promoter recognition and genome internalization by the newly synthesized T7 RNAP. The slow entry of the T7 genome into the cell provides an obvious mechanism for T7 gene regulation, as early genes are the first, and late genes the last, to enter the infected cell. However, patterns of gene expression and phage yields are fairly normal, albeit a few minutes faster than usual, when T7 DNA is rapidly introduced into cells from a λ virion (82). How this is achieved is not understood, but the presence of the strong late T7 promoters in the cell does not prevent the weaker class II promoters from functioning.

Gene expression and phage yield are also fairly robust to a significant change in gene order (65). The major biological reason for coupling transcription to genome internalization is to ensure that the first gene product to be made, gp0.3—an antirestriction protein—functions before type I enzymes recognize their cognate sites on the entering DNA. T7 is normally resistant to type I restriction but becomes sensitive when its genome enters the cell from a λ virion (82).

T3 is also known to internalize its genome via transcription (208, 324). The close genetic similarity of the *Yersinia* phages ϕA1122 and ϕYeO3-12, and perhaps phage SP6 and the *Pseudomonas* phage gh-1, to T7 and T3 implies that their genomes are also internalized by transcription. However, *P. aeruginosa* RNAP cannot internalize the ϕKMV genome by a comparable process. The time taken for host RNAP to transcribe the phage RNAP from near the genome end is about the entire latent period. Thus host RNAP-catalyzed genome internalization is unlikely to be a distinguishing feature of the T7 family.

Early Gene Expression

The three A promoters near the left genome end are among the strongest known and direct essentially all transcription by *E. coli* RNAP in vivo. In their absence the minor B or C promoters can be used to drive transcription-mediated genome entry and synthesis of the gene *1* mRNA. Several other *E. coli* σ^{70} promoters have been found on the genome by in vitro studies, or inferred from the properties of plasmids containing T7 DNA. None seem likely to have significance in vivo, however. Antitermination of transcripts at a *boxA* element between the A promoters and the first gene, *0.3*, is likely the reason why no amber mutants exhibiting transcriptional polarity have been found in T7. The terminator TE downstream of gene *1.3* stops most transcribing complexes and thus defines the end of the class I, or early region of the genome (123, 270). Termination at T7 TE is Rho-independent both in vitro (28) and in vivo but its efficiency is dependent on phosphorylation of RNAP by the gp0.7 protein kinase (211, 330). Interestingly, T3 TE, which is almost identical by sequence to T7 TE, terminates transcription efficiently in vivo but not in vitro. A novel transcription factor, tau, is required for termination at T3 TE and it also enhances termination at T7 TE (15, 16). Transcription that reads through TE terminates specifically (but also inefficiently) near the *3.8* promoter and RNase III recognition site, and at Tϕ, the terminator for T7 RNAP. *E. coli* RNase III processes the approximately 7 kb early RNA at five sites. Processing upregulates expression of gene *0.3* (61) but has little effect on other genes (early or late) and phage growth is almost normal in RNase III (*rnc*) null mutants.

As a consequence of gene *0.7* activity, host RNA and protein synthesis have stopped by about 4 minutes of infection at 30°C, and as a consequence early T7 gene expression has ceased by about 8 minutes (173). T7 RNAs have been reported to be stable (166, 282), or chemically stable but functionally unstable (315–317). Competition for ribosomes of the *0.3* mRNA by the more abundant late mRNAs was proposed to explain the shutoff of gp0.3 synthesis (264–266) but discrimination between early and late mRNAs was not observed in another study (312). Gene *0.7* positively regulates RNase III activity (172), which should stabilize mRNAs; furthermore, gene *0.7* also inhibits RNase E (165), the major mRNA degrading activity in *E. coli*. The issue of T7 mRNA stability and translational discrimination clearly warrants further investigation.

Early Proteins

The functions of five early proteins are known in detail. Gene *0.3* is expressed at very high levels, and protects T7 from various type I restriction systems (2, 3, 273). Gp0.3 has been crystallized (301); the protein is a dimer and mimics the shape of B DNA. Gp0.3 has been shown to bind the *Eco*KI restriction enzyme in a 2:1 complex and to displace the enzyme from its target DNA (2). Although T7 gp0.3 shares little sequence similarity with T3 gp0.3, the latter not only provides the same activity as T7 gp0.3 but also hydrolyzes AdoMet, the methyl donor for DNA modification (278). Both the T7 and T3 types of gene *0.3* are found in other T7 group members.

Gene *0.7* codes for a protein that inactivates host-catalyzed transcription. It is also a serine-threonine protein kinase (216). The two activities are separable (253). Early T7 RNA and protein synthesis is prolonged after infection by *0.7* mutants. Gp0.7 phosphorylates many host proteins, including ribosome components (220, 221), ribonucleases III and E (165, 172), and the β and β′ subunits of *E. coli* RNAP (330). The BR3 (271) and Y49 strains (98) contain the *rpoC-E1158K* mutation (200), which renders *0.7* essential for phage growth. The mutation affects the region of β′ to which the T7 protein gp2 binds. Gene *0.7* is nonessential in wild-type hosts and is actually detrimental when T7 is grown at 30°C in rich media. However, phage growth at elevated temperatures or in minimal media, or in cells harboring the Col Ib plasmid, renders gene *0.7* essential (84, 102). The processes underlying the conditional necessity for *0.7* function have not been determined.

Gene *1.2* inhibits *E. coli* dGTPase and converts it into an rGTP-binding protein (5, 105, 198), which is important for T7 only when only the dGTPase (*dgt*) gene is overexpressed (*optA1* mutation). T7 gene *1.2* also causes phage exclusion by the F plasmid, but T3 gene *1.2* overcomes the exclusion system (188). The ATP-dependent DNA ligase (gene *1.3)* is only essential in *E. coli lig* mutants (271), but its presence enhances phage growth in wild-type hosts. The protein was the first of the nucleotidyl transferase superfamily to be crystallized (280).

T7 RNA polymerase

T7 RNAP (gene *1*) has been crystallized (257); structures have also been obtained with gp3.5 lysozyme, and with DNA in both initiation and elongation complexes (30, 31, 114, 287, 290, 320, 321). T7 RNAP is a single-chain enzyme of 883 amino acids that is related by sequence to mitochondrial and chloroplast RNAPs, and by structural considerations to the DNA polymerase I family. A Y693F change in T7 RNAP allows efficient usage of dNTPs as substrates (258). T7 RNAP and its interactions with both DNA and RNA have also been intensively studied by mutational and crosslinking approaches and by kinetic analyses.

The 23 bp T7 promoter can be divided into a binding region, which extends from −17 to −5 and includes the determinants for promoter specificity, and an initiation region that extends from −4 to +6 (111). Both values are relative to the transcription start site at +1. T7 RNAP binds to one face of the closed DNA duplex promoter (294, 295) via the specificity loop (amino acids 739–770) (30) and an AT-rich recognition loop (amino acids 93–101). Promoter opening between base pairs −5 and −4 is thermodynamically driven (4), being favored by promoter bending (296) and by a β-hairpin of RNAP that includes V237 (260). Initiation of transcription is preferentially with GTP (167) and is very fast, rates of approximately 3 per second having been observed in vivo (156). The unstable initiation complex is converted into a stable elongation complex when the nascent RNA has reached approximately 12–14 nucleotides in length (179). Conversion involves an interaction of the specificity loop (289) and is accompanied by a major change in enzyme conformation (110, 160, 168, 259). This can be seen by comparison of the structures of RNAP as initiation and elongation complexes (30, 287, 320). The rate of mRNA synthesis catalyzed by T7 RNAP is at least 5-fold faster than the *E. coli* enzyme and has been estimated to be 200–240 bases per second at 37°C, both in vitro and in vivo (14, 163, 329).

Although T7 RNAP is highly specific for its promoter, natural promoters vary in sequence. The 17 promoters on the T7 genome can be divided into three groups, six consensus class III promoters (including ϕ*OR*), 10 class II promoters that vary from consensus by two or more positions, and the promoters close to each genome end, designated ϕ*OL* and ϕ*OR*. ϕ*OL* and ϕ*OR* are not known to be important for gene expression and have been described as replication promoters. Both ϕ*OL* and ϕ*OR* have been reported to initiate replication in mutants lacking the primary origin (59, 215, 288).

RNAPs of different members of the T7 family may recognize different promoters, and six specificities are currently known. As more distant relatives of the family are characterized, this number will likely increase. There is surprisingly little cross-talk between RNAP and heterologous promoters even though a single amino acid change in T7 RNAP is sufficient to switch specificity to that of a T3 promoter (218, 219, 226). Similarly, a single change at −11 in the 23 bp promoter is sufficient to cause a specificity switch (131); two changes in the T7 promoter also allow recognition by phage SP6 RNAP (152). A comprehensive study on SP6 promoter variants has been completed (249).

The activity of T7 RNAP is modulated by gp3.5 lysozyme. Mutants that exhibit altered susceptibility to lysozyme have been characterized (186, 328). It is not obvious from the crystal structure of the RNAP–lysozyme complex (114) how all the mutations, which are scattered across the RNAP gene and alter widely spaced residues, exert their effect. However, the large conformational changes that occur in the enzyme during the transcription cycle are not captured in the structure of the RNAP–lysozyme complex. Lysozyme is responsible for the switch from class II to class III gene expression (109, 174, 175, 186) during infection. Lysozyme destabilizes initial transcription complexes (104, 139, 300, 329), and as T7 class II promoters are inherently weaker than their class III counterparts the former are preferentially affected. Although remaining bound to RNAP, lysozyme has no effect on an elongation complex. However, transcription termination may be affected.

T7 RNAP is relatively insensitive to *E. coli* RNAP termination signals, although some, including *rrnB* T1 and T2, and λ tR′, are moderately efficient (34, 91, 113, 158, 159, 161, 162, 329). Notably, TE, which terminates transcription by *E. coli* RNAP at the end of the early region, has no effect on T7 RNAP. Two T7 RNAP terminators exist in the phage genome. The major late terminator, Tϕ, is located distal to gene *10* and consists of a long stem-loop followed by a string of six U residues. T7 Tϕ is not totally efficient, and read-through is actually required as there is no promoter between the terminator and the essential genes *11* and *12*. Termination efficiency at Tϕ is not affected by lysozyme (329). In contrast, lysozyme is absolutely required at the second terminator, CJ, named for its location near the concatemer junction. On the normal genetic map of T7 CJ lies immediately downstream of the left terminal repeat. CJ has the sequence 5′-ATCTGTT that has no obvious secondary structure and is not followed by a string of U residues. In the absence of lysozyme CJ acts primarily as a pause site but does not cause termination; the presence of lysozyme greatly enhances pausing, which results in substantial levels of termination (158, 159, 329). Termination is thought to be due to the paused elongation complex returning to an unstable initiation complex (259), which is then destabilized by lysozyme (104, 300). Pausing or termination of transcription at CJ reflects the role of lysozyme in stimulating DNA replication and packaging of DNA during infection (329). The CJ heptamer sequence is conserved in the closest relatives of T7 but not in phages more distantly related.

Late Gene Expression

Class II genes (*1.1–6.3*) are expressed from about 6 minutes through 15 minutes of infection at 30°C. The first three genes, *1.1–1.3*, in the class II region are thus transcribed early by *E. coli* and late by T7 RNAP. Ten promoters direct expression of T7 class II genes, but this organization is not conserved in all T7-like phages. Most class II (and some class III) RNAs terminate at Tϕ, providing overlapping transcripts that have different 5′ ends. Thus, genes lying at the distal end of the class II region are transcribed more frequently than those near the beginning. Relative to promoters in the T7 class III region, class II promoters are weak but this hierarchy is not maintained in all T7-like phages. For example, the sequences of the T3 class II promoter *ϕ2.5* and the class III promoter *ϕ9* are identical (7). However, RNA concentrations may not be the major regulator of gene expression, and different translational efficiencies may account for most variations in the levels of class II proteins synthesized. For example, gp2.5 and gp3.5 are among the most actively synthesized proteins by T7, but gp2.8 and gp3, which are translated from the same set of overlapping RNAs, are not abundant proteins. Almost all T7 coding regions are preceded by "good" Shine–Dalgarno sequences; translational coupling (225, 232) and the use of GCU as the second codon for highly expressed genes (62, 83) are two factors governing levels of proteins made during infection.

Transcription of class II genes is controlled by the class II protein lysozyme (gp3.5) by a feedback mechanism. Gene *3.5* mutants show extended transcription of class II genes, and reduced levels of both DNA replication and DNA packaging. Reduced replication could be due to high levels of transcription from the *ϕ1.1A* and *ϕ1.1B* promoters, which synthesize the initial primer RNAs, but this has not been demonstrated experimentally. Reduced DNA packaging is due to the failure of the CJ pause/terminator signal to function in the absence of lysozyme. Although transcription of class II genes is shut off by lysozyme, the question remains why class II protein synthesis also stops. Competition for ribosomes between class II mRNAs and the more abundant class III mRNAs is again a possibility, but the functional stability of class II mRNAs remains to be studied carefully.

Class III promoter sequences are totally conserved; transcripts from *ϕ6.5*, *ϕ9*, and *ϕ10* inefficiently terminate at Tϕ; read-through RNAs and transcripts from *ϕ13*, *ϕ17*, and *ϕOR* terminate at the genome end or, if concatemeric DNA is the substrate, perhaps at CJ. The gene *17* promoter cannot be of major importance as T3 *ϕ17* has T7 specificity and is not used during infection (228). In contrast to many other phages, in T7 late transcription is not coupled to DNA replication. In fact, incorporation studies of ^{32}P-phosphate into T7 RNAs (8) suggest that transcription may be largely over before much replication has occurred. Furthermore, late gene expression appears normal when DNA replication is totally prevented by mutation (270).

Class II Proteins

Gp2 binds to and is an inhibitor of *E. coli* RNAP (97–99, 200), but in K-12 or B strains the defects due to the lack of gene *2* activity are reduced phage DNA replication and premature breakdown of replicating phage DNA, specifically at the left genome end where the major *E. coli* promoters are found (55, 192, 313, 314). Breakdown of replicating T7 requires both gene *3* endonuclease and DNA packaging activities. The gene *2* protein has been purified both by its inhibitory activity (54, 99) and by its ability to stimulate DNA packaging in gene *2* mutant-infected cells (146). After early transcription has been completed a 2^- infection can be rescued by addition of rifampicin during infection (192, 204), and the drug also substitutes for gp2 in DNA packaging assays (146). Growth on *E. coli* B or K-12 *tsnB* strains requires that T7 carries a gene *2* mutation(s) although T3$^+$ growth is normal (238). The *tsnB* mutation is now known to be *rpoC-E1188K* (200); the change in RpoC prevents inhibition by wild-type gp2 but not by the T7 gene *2* mutant β or T3 gp2. The ability of T7 gene *2* null mutants to grow in *E. coli* C does not, however, appear to be due to differences in RNAP as transductants of *E. coli* C containing RNAP subunits from *E. coli* K-12 strains remain permissive (G. E. Christie and I. J. Molineux, unpublished observations). An *E. coli* K-12 mutant that is deleted for the region of RpoC that binds gp2 is nonpermissive for T7 (63; W. P. Robins and I. J. Molineux, unpublished observations). T7 mutants selected to grow on this host include some that affect the early *E. coli* promoters, but they have not been analyzed in detail.

The defects due to the inability of gp2 to inhibit *E. coli* RNAP occur late in infection. Host transcription is normally inactivated by T7 gp0.7 before DNA replication and packaging occur, and neither gp2 nor rifampicin can remove promoter-bound *E. coli* RNAP. It is unclear how *E. coli* RNAP interferes with the process of DNA maturation and packaging. Nevertheless, both in vitro (308) and in vivo (55) the absence of gp2 leads to selective degradation of the left end of the T7 genome. It is obvious that there is much yet to understand about the biological roles of gp2 and *E. coli* RNAP during T7 infection.

Gene *2.5* encodes a single-stranded DNA-binding protein (SSB), which is essential for T7 DNA replication and recombination (128). The structure of the protein has been determined to 0.19 nm (103); it exists as a dimer that, unlike *E. coli* SSB or T4 gp32, binds to DNA with little or no cooperativity (129, 130). However, T7 SSB is much more effective at mediating homologous base-pairing than its *E. coli* or T4 counterparts. SSB interacts with T7 primase-helicase, mediating efficient strand exchange and stimulating primase activity (133, 134). SSB also interacts with T7 DNAP and is therefore the primary organizer for coordinating leading- and lagging-strand synthesis (130, 148, 149).

Genes *3* and *6* code, respectively, for an endonuclease that prefers ssDNA and a 5′→3′ exonuclease; both enzymes are essential for recombination in vivo (150, 151). Both proteins are also required for degradation of the host chromosome during infection and their combined action provides the nucleotides for subsequent phage DNA replication. About 80% of the DNA in progeny phage is derived from the host chromosome. Both gp3 and gp6 can degrade T7 DNA in vitro; how their activities are controlled in vivo to prevent destruction of the infecting genome is unknown. The gene *3* endonuclease is also a Holliday structure resolvase (50, 51, 157) whose crystal structure is known (86). The gene *6* exonuclease also possesses an RNase H activity that removes RNA primers (250, 251) and promotes concatemer formation during DNA replication (147).

Gene *3.5* codes for lysozyme, which is actually not a lysozyme but an *N*-acetylmuramyl-L-alanine amidase (112). The most important functions of lysozyme are also not in cell lysis but in controlling initiation and termination by T7 RNAP, which in turn affect DNA replication and packaging (329). These properties better reflect its position among the class II DNA metabolism functions rather than with the class III lysis-associated genes *17.5* and *18.5*/*18.7*. Lysozyme is, however, important in releasing progeny phage from the debris of lysed cells (252, 327). The nucleic acid metabolism and amidase functions of lysozyme can be separated. The N-terminal residues of the protein are essential for interaction with T7 RNAP, removing residues 2–5 prevents binding, and causes the same defects in transcription, replication, and packaging as a complete gene deletion (33, 327). The deletion leaves amidase activity intact. In contrast, removing residues 130–135 (of the 150 amino acids in the mature protein) inactivates amidase activity but allows a normal interaction with RNAP. Gene *3.5* amber mutants make small plaques at an efficiency of approximately 0.5 on nonsuppressing hosts, whereas a *3.5* deletion mutant makes pinpoint plaques at an efficiency of approximately 0.2, although about half the normal amount of replication occurs and the burst size is about one third that of wild-type (327). Addition of T7 lysozyme (or detergent) to the lysed culture is necessary to complete release of progeny phage, which appear to be trapped. Although cell lysis by *3.5* mutants is delayed and incomplete, it is not markedly different from that after infection by mutants blocking DNA replication (252, 269). The timing and normal abruptness of lysis of T7-infected cells may be determined in part by DNA replication and/or packaging events.

The full-length gene *4* protein is both a 5′→3′ helicase of the *E. coli* DnaB family *and* a DNA primase. Both enzyme activities have been the subject of recent reviews (69, 255). An in-frame internal start site within gene *4* results in a protein, gp4B, which lacks the N-terminal 63 amino acids of the 63 kDa gp4A and consequently lacks primase activity (10, 11). Both proteins are made in infected cells; the native form of the enzyme appears to be a heterohexameric ring but each protein alone makes a similar structure (64, 203). The enzyme also forms heptameric rings (291). Single-stranded DNA passes through the central hole of the ring (322). Crystal structures of the separate primase and helicase domains, and of the intact protein, have been reported (116, 254, 291). Both the crystal structure of the intact protein and single-particle electron microscopic image reconstructions (299) show that the primase active site lies on the outside of the hexameric ring. Although most closely related members of the T7 family make two gene *4* proteins, the phage SP6 primase gene lacks an internal start; furthermore, helicase activity of the protein was not demonstrated (292).

The N-terminal region of gp4A contains a Cys4 zinc finger motif that makes sequence-specific interactions with a template for RNA primer synthesis (67). The primase recognition sites on single-stranded DNA are 5′-G/TGGTC, and 5′-GTGTC; the 3′ base is necessary for recognition but is not copied into primers, which thus have the sequence 5′-ACCA/C and 5′-ACAC (68, 70, 178, 286). Helicase hydrolyzes most NTPs, with dTTP being most efficient, and hydrolysis is stimulated by single-stranded DNA (169, 170, 209, 305). Hydrolysis of dTTP by the hexameric helicase has been likened (101) to the binding change mechanism of the rotary F_1-F_0 ATPase (202). The rate of translocation of helicase along single-stranded DNA has been measured at 18°C to be 132 bases per second (125), comparable to the rate of in vitro DNA replication.

T7 DNAP consists of two proteins: the gene *5* protein and the processivity factor *E. coli* thioredoxin. Thioredoxin forms a 1:1 complex with the T7 protein and increases its processivity by at least 1000-fold. The complex also increases the activity of the associated 5′→3′ double-stranded DNA exonuclease but has little effect on single-stranded DNA exonucleolytic activity (107, 285). The redox function of thioredoxin is not important in complex formation (106), rather the protein plays the equivalent function to that of sliding clamps in other replication systems. A crystal structure of T7 DNAP bound to a primer-template and with a dNTP in the active site has been obtained at 0.22 nm resolution (58). Thioredoxin, and the loop on gp5 to which it binds, may encircle the primer-template DNA exiting the polymerase. T7 DNAP, like the RNAP, is a member of the DNA polymerase I family with palm, fingers, and thumb domains. Thioredoxin binds to an extended 71 residue loop in the thumb domain; interestingly, this loop is not present in *E. coli* Pol I. Pol I synthesizes DNA distributively but engineering the thioredoxin-binding sequences into the appropriate location confers thioredoxin-dependent processivity to the enzyme (9).

Gene 5 protein, thioredoxin, gp4 primase-helicase, and gp2.5 SSB together catalyze simultaneous and coordinate replication from a synthetic minicircle double-stranded DNA template with two primase recognition sites and a single-stranded 5′ tail (148, 149). This cleverly designed reaction synthesizes DNA at approximately 300 bases per second per polymerase molecule at 30°C, close to the rate observed in vivo. When sufficient SSB is provided, both leading and lagging strands are synthesized at the same rate, and more than 30kb duplex DNA molecules are produced.

Replication In Vivo

T7 DNA replication is independent of host DNA replication, repair, and recombination functions. Although eroding one strand of each terminal repeat allows formation of circular T7 molecules in vitro, such molecules are not found in vivo. Nevertheless, T7 replication is extremely sensitive to the topoisomerase inhibitor nalidixic acid (136). By trapping covalently bound topoisomerase on DNA, nalidixic acid inhibits T7 DNA replication. However, growth of T7 is normal in *gyrA*(Ts) mutants at their nonpermissive temperature. Similarly, normal T7 growth occurs in *gyrA*(Ts), *gyrB*(Ts), *gyrA*(Ts)*gyrB*(Ts) double mutants, *par C*(Ts), and *parD*(Ts) mutants at high temperatures (P. Kemp and I. J. Molineux, unpublished observations). Furthermore, at the non-permissive temperature for DNA gyrase or topoisomerase IV, T7 growth is fully resistant to quinolone antibiotics (P. Kemp and I. J. Molineux, unpublished observations; 136).

The primary origin of replication is an AT-rich region lying immediately downstream of the *ϕ1.1A* and *ϕ1.1B* promoters. Both are used by T7 RNAP to initiate replication on the leading strand. In vitro, primers of 10 to 60 bases originate from both promoters (76, 77), whereas in vivo the transition from RNA to DNA occurs at many sites over a few hundred base pairs (281). In the absence of T7 RNAP, no DNA replication occurs, but as RNAP is also necessary for late gene expression few, if any, replication proteins are made in its absence. Replication of DNA absolutely requires the presence of genes *1*, *2.5*, *5* (and *E. coli trxA*), and gene *4A*. Mutants lacking the helicase-only gene *4B* grow normally (177, 230). In vitro, gp2.5 has been shown to be necessary for bidirectional replication (77). Mutants lacking genes *3* or *6* show a premature replication shutoff; the host chromosome is not degraded and the nucleotide precursors required for replication are therefore not produced. In *optA1* cells, which overexpress a dGTPase, the lack of gene *1.2*, which inhibits this host enzyme, also leads to a deficit of nucleotide precursors and the consequent premature inhibition of replication (5, 105, 311). T7 DNA is degraded by genes *3* and *6* in this abortive infection (233).

T7 replication is bidirectional, but the overall movement of the left fork is delayed relative to the right fork. Thus the origin of replication determined by electron microscopy (where the midpoint of a replication bubble is measured) is displaced from the true origin. Deletion mutants that lack *ori* grow well, perhaps even better than wild-type under laboratory conditions. Thymidine-labeling of newly synthesized DNA suggested that replication in these mutants initiates from secondary origins associated with the *ϕ6.5*, *ϕ13*, and *ϕOR* promoters (215). Replication from *ϕOL* was not detected in this study. In contrast, density fractionation of light T7 DNA that was replicating in heavy (^{15}N, ^{2}H) media, followed by electron microscopic analysis, showed the most pronounced secondary origin to be near the left genome end (59, 288). This origin is possibly associated with *ϕOL*. The difference in the two studies may reflect the different growth conditions employed, which in turn suggests that different replication origins may be used preferentially by T7 as it grows in hosts that are in different physiological states.

Plasmids containing *ϕOR* (and certain other promoters) are replicated extensively in T7-infected cells (36, 37, 214) and can inhibit phage growth (20, 36). Plasmids that contain both *ϕOR* and the terminal repeat TR region are not only replicated but also packaged to form transducing particles (36, 37). Plasmids containing the comparable regions of T3 respond similarly after T3 infection (293).

T7 DNA is linear; consequently, the first round of replication produces two incomplete duplex molecules with 3′ single-stranded extensions. These extensions are part of the terminal repeats and can thus anneal to form a linear concatemer; the latter thus contains a single copy of TR between two genomes. Concatemer formation is known to require the gene *6* exonuclease in vivo (147) and may also require SSB to promote strand annealing. Linear concatemers containing up to about 10 genome equivalents are found early after infection but then a fast-sedimenting complex appears that contains >100 genome equivalents (205). This complex is not torsionally constrained, it contains many single-stranded regions, and it consists of genomes in the process of both recombination and replication (142, 143). T7 recombination is very efficient. Recombination intermediates are likely converted into replication forks, similar to what is found in T4 DNA replication. Consistent with this idea, the bulk of progeny T7 DNA is synthesized within about 5 minutes at 30°C (the more extended period seen by pulse-labeling with thymidine reflects a reduction in nucleotide pool size, and a time-course of thymidine incorporation is therefore not an accurate indicator of DNA synthesis rates throughout infection). The complex DNA structure is eventually resolved by the Holliday structure-resolving gp3 endonuclease back to linear concatemers (51, 157), which are the substrates for DNA packaging.

Class III Proteins

Class III proteins are all involved in virion assembly, DNA packaging, or cell lysis. The latter process has not been studied in detail. In contrast to most phages, T7 lysozyme is not required for lysis as measured by a decrease in culture turbidity, although it is important for releasing progeny from cell debris. Lysis is delayed after infection by a gene *3.5* null mutant but no more so than by any mutant defective in DNA replication (269, 270, 327), and lysis may actually be triggered by some aspect of DNA replication or packaging. Gene *17.5* codes for a typical type II holin (302) but *17.5* amber mutants of both T7 and T3 still lyse, albeit less synchronously and with a delay relative to wild-type. Whether the amber mutations employed were leaky or whether T7 has an alternative means of disrupting the cell membrane is not known. Note that T7 does not contain a typical lysis "cassette" of adjacent holin and lysozyme functions. In addition, genes *18.5* and *18.7* are separated from holin by the small terminase subunit. Genes *18.5* and *18.7* are homologs of λ *Rz* and *Rz1*, and were shown to complement P22 gene *15* mutants for lysis in the presence of divalent ions (21). T7 *18.5* or *18.7* mutants have not been tested for a lysis defect. Interestingly, although close relatives have the same genetic structure as T7, the *P. aeruginosa* phage ϕKMV genome has been annotated with the more traditional lysis cassette (144).

Virion Assembly

The head assembly pathways of T7, T3, ϕII, and probably most close relatives of T7 are the same (246). Proheads contain the head–tail connector gp8, scaffolding protein gp9, capsid protein gp10, and the three internal core proteins gp14, gp15, and gp16 (24, 26, 223, 241, 244). Early reports on the functions of genes *7* and *13* in assembly should be largely discounted. Gene *7* was originally defined by mutations that mapped between genes *6* and *8* (269) but was subsequently shown to be two genes, *7* and *7.3* (62). Both are now known to be mature virion proteins, but gp7 is nonessential whereas gp7.3 is a tail protein that is essential for infectivity although not for particle formation (P. Kemp and I. J. Molineux, unpublished observations).

In the absence of gp9 or gp10, no capsid-like structures form (223, 244), whereas in the absence either of the connector or the three internal core proteins aberrant polycapsids appear (24, 223, 241, 242, 261). All six proteins are required for prohead formation in vivo but the order of their assembly remains unclear. The dodecameric connector and the three core proteins may assemble and serve as an initiator for scaffold and capsid protein assembly; alternatively, the scaffold and capsid protein may assemble to form incomplete prohead shells, which are then closed by insertion of a preformed connector-core complex (24). After DNA packaging the tail proteins cooperatively assemble on the filled head. No tail structures independent of a head are found (171, 277).

DNA Packaging

T7 packaging has been examined in vitro and in vivo, although studies on T3 have been more comprehensive. A defined in vitro system, consisting of purified proheads and the terminase proteins gp18 and gp19, packages unit-length DNA. ATP is used as energy source, both for DNA translocation and DNA condensation inside the prohead (88).

A crude lysate of T3-infected cells was shown to package T3 DNA in vitro (72) and to discriminate between T7 and T3 DNA substrates. Unit-length genomic DNA was active in this assay but was converted into concatemers during the reaction (75). Discrimination is conferred primarily by the specificity of the phage RNAP for the promoter ϕ*OR*, for downstream DNA, and by sequences at the terminal repeat TR necessary for the initial cleavage of concatemeric DNA (92, 93, 318). Comparable bipartite DNA sequences required for packaging are found in T7 (37, 38). A more purified in vitro packaging system, consisting of proheads and the gp18 and gp19 terminase, packages any linear double-stranded DNA without prior conversion to concatemers (88).

The initial events of packaging T3 DNA are comparable to those of other double-stranded DNA phages (73, chapter 6). DNA is recognized by the small terminase subunit gp18, while the large subunit gp19 binds the prohead. There is genetic evidence for a role of the connector protein in T3 DNA packaging (71, 74, 199) and the T7 connector has also been shown to bind DNA (25). Using ATP as an allosteric effector, a 50S ternary complex forms between the T3 prohead, DNA, and terminase; packaging is then initiated by ATP hydrolysis (247, 248). The large terminase subunit contains the ATP hydrolysis, the DNA translocation, and both the specific and nonspecific dsDNA endonucleolytic activities (193, 194, 195). Specific cleavage is not required for packaging unit-length DNA but is of course necessary to make a mature genome from concatemeric DNA. ATP hydrolysis is coupled to DNA translocation into the prohead. Each ATP molecules packages approximately 1.8 bp at an average rate of 22 kb per minute at 30°C (193, 248). T7 DNA is packaged cooperatively from concatemeric DNA and at a comparable rate to that of T3 DNA in vitro (256, 283, 284).

"Solutions" to the problem of completely replicating the linear T7 DNA molecule have existed for over 30 years (117, 307). These models have long been known to be incorrect in detail. How the single copy terminal repeat (TR) between genomes in the concatemer junction is duplicated in producing mature progeny genomes is still not fully understood. The failure to duplicate the TR necessarily results in only half the replicated DNA being packaged into viable progeny particles. Although this would seem to be a very inefficient use of genetic material, it should be noted that normal lysates do contain substantial amounts of unpackaged phage DNA. Nevertheless, it is generally supposed that T7 does have a mechanism to duplicate the TR during maturation and packaging. In vitro studies suggest that the two mature ends of T7 DNA are formed in uncoupled reactions that involve double-stranded DNA cleavages (308). The initial double-stranded DNA break is made at the right end of the concatemer junction (corresponding to the genome right end) by the gp18–gp19 terminase. In vivo the mature right end is known to be formed prior to the mature left end (245), but how the genome left end is formed is less clear. A covalently closed hairpin containing the TR is found during T7 infection (38), and is also found in cells replicating plasmids that contain the concatemer junction (36, 37). Sequencing the hairpin

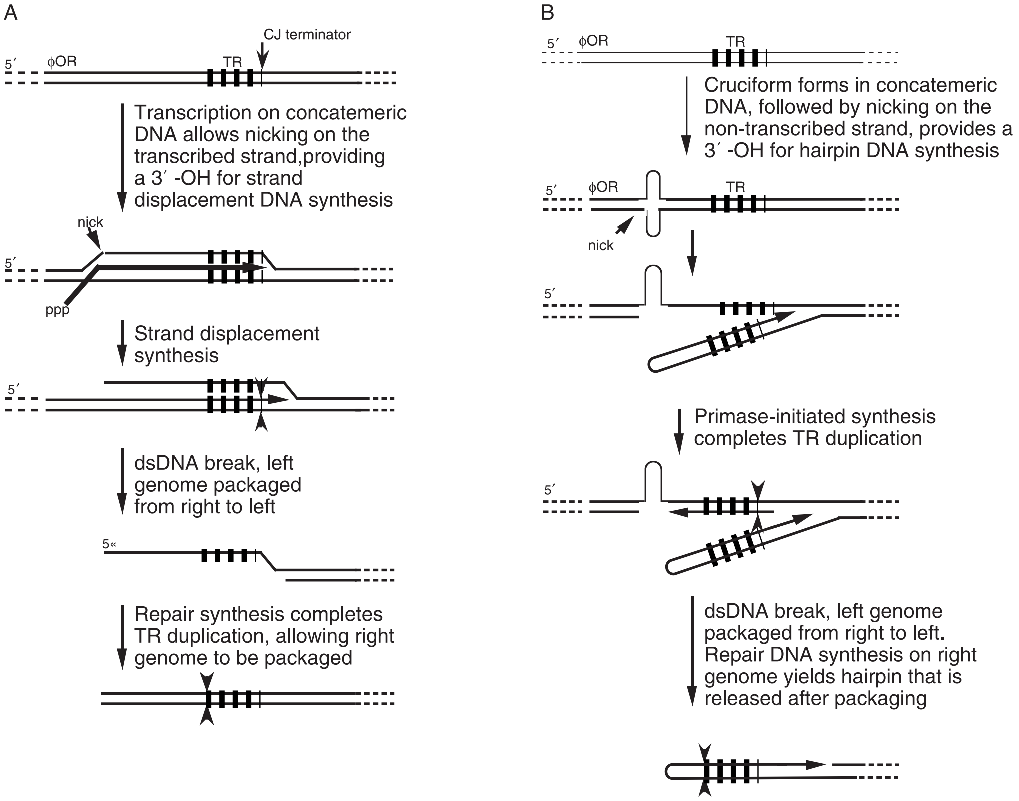

A
CJ terminator
5′
ϕOR
TR
Transcription on concatemeric DNA allows nicking on the transcribed strand,providing a 3′ -OH for strand displacement DNA synthesis
nick
ppp
Strand displacement synthesis
dsDNA break, left genome packaged from right to left
Repair synthesis completes TR duplication, allowing right genome to be packaged
B
Cruciform forms in concatemeric DNA, followed by nicking on the non-transcribed strand, provides a 3′ -OH for hairpin DNA synthesis
Primase-initiated synthesis completes TR duplication
dsDNA break, left genome packaged from right to left. Repair DNA synthesis on right genome yields hairpin that is released after packaging

Figure 20-3 Proposed mechanisms for TR duplication. Adapted from Fujisawa and Morita (73) and Chung and Hinkle (37).

structure DNA suggested the model shown in figure 20-3B for duplication of the terminal repeat (38). The nuclease that nicks the cruciform to generate the hairpin is known not to be the gene *3* product, but has not yet been identified. It was suggested that it may be the gene *19.5* product, but the latter is not essential for the formation of transducing particles (293), and neither *19.5* nor the cruciform region are essential for phage growth (126, 127). The most pronounced defect due to loss of *19.5* is a delayed breakdown of host DNA and a reduced burst. Deleting the cruciform DNA causes a lysis delay.

As originally described, the hairpin model does not account for the role of transcription termination at the CJ pause/terminator signal found immediately downstream of TR. The failure of a T7 RNAP mutant that is transcriptionally competent but pause/termination-defective at CJ to support DNA packaging can be suppressed by a gene *19* mutation (158, 328). This observation led to the idea that the paused or terminating RNAP at CJ allows the nontemplate strand to be nicked by terminase. Strand displacement synthesis then allows duplication of the TR sequences (73, figure 20-3A). However, this model does not explain the existence of hairpin DNA, which is thought to be an intermediate in formation of the mature left genome end. Nor does it satisfy the biochemical observation that terminase causes a double-strand break. The difficulty in establishing the mechanism of TR duplication suggests either that more than one efficient pathway exists or perhaps that TR duplication is not necessary for a substantial burst of progeny phage.

Exclusion of T7 by the F Plasmid

T7 growth is aborted in cells containing the F plasmid (164). The only F function required is *pifA* (46, 47, 235, 236), a gene that is nonessential for conjugation or plasmid maintenance. A similar gene is found in a few other conjugative plasmids, including R56 and R64 (108,180, 306). Proteins closely related to PifA are also predicted in a pathogenicity island of a uropathogenic *E. coli* strain and in a cryptic prophage of *S. typhimurium*. More distant relatives are found in both Gram-negative and Gram-positive organisms, and patches of the PifA amino acid sequence show similarity to a number of ABC transporters and other ATP-binding proteins. Attempts to demonstrate ATP binding to F PifA have not been successful. Other than inhibiting growth of T7 and related phages, PifA has no defined function.

The *pif* operon contains only *pifA* and an autoregulated repressor *pifC* (46, 47, 181, 182). Early reports of a *pifB* gene are incorrect. PifA is a membrane-associated protein but has not been shown to cross the cytoplasmic membrane (32). PifA interferes with the normal function of both gene *1.2* and gene *10*, and the interaction of PifA and either phage protein results in the inhibition of all macromolecular synthesis in infected cells. Inhibition requires the accumulation of T7 gp1.2 or gp10 but the slow penetration of T7 DNA into the infected cell means that the infection is aborted before the complete genomes of about half the infecting phages have entered the cell (80). Early T7 gene expression is therefore normal in cells containing F, but expression of genes closer to the right end of the genome is severely affected (8, 12, 13, 40, 196, 197). Exclusion of T7 also results in the loss of membrane integrity. Membrane transport processes are inhibited, and intracellular ATP and other phosphorylated molecules leak from the abortively infected cell (41, 234, 235, 236, 237). A major problem in understanding the process of exclusion is in distinguishing the initial event that then leads to a plethora of physiological dysfunctions (188).

Mutants of T7 that grow in cells containing F have been isolated (190). They necessarily contain a missense or null mutation in gene *1.2* and two missense mutations in gene *10*. Although T3$^+$ grows normally in F cells, T3 *1.2* mutants behave like wild-type T7 and are excluded (191). In the *1.2* mutant background two missense mutations in T3 gene *10* are required to restore growth (42). T3 gene *1.2* normally therefore protects the PifA-containing cell from expression of gene *10*. T7 gene *1.2* is not only unable to perform this function but itself causes the infection to abort.

The physiological defects associated with F exclusion of T7 can be mimicked by using cloned genes in the absence of phage infection. Any plasmid expressing *pifA* is lethal to a cell that also expresses a cloned gene *1.2* or gene *10*. Lethality is completely suppressed when the cloned gene *1.2* contains a missense mutation first recognized in mutant phages that grow in F cells, but is only reduced with comparable mutant gene *10* plasmids (237). Several independent selections for *E. coli* mutants that tolerate the combination of *pifA* and gene *1.2* or *pifA* and gene *10* resulted in the isolation of a single mutant (303). The mutation is a promoter-up that increases expression of *fxsA* about 25-fold. A disruption of *fxsA* still leads to F exclusion of T7, thus FxsA is not the cellular target of PifA and either of the T7 proteins. Overexpression of *fxsA* therefore provides a protective function that alleviates most of the physiological defects associated with F exclusion of T7 and allows a normal phage burst (304). *fxsA* overexpression seems to affect the mechanism of exclusion directly, which is different from most other chromosomal mutations (e.g., *strA*, *rif*, *rho*, *gyrA*, *galU*, *himA*) that affect the exclusion phenotype by affecting expression levels of *pifA* or overall cell physiology (121, 236).

FxsA is a membrane protein of unknown function with four transmembrane segments (32, 304) and a C-terminal cytoplasmic tail. This tail is not important for overcoming exclusion, but the fourth transmembrane segment is essential (32). The membrane association of PifA and the protection from lethality by the integral membrane protein FxsA strongly suggest that the cell membrane is the site of the initial events in the exclusion process.

PifA interacts with gp1.2 and with gp10 in vitro, and all three interact with FxsA, but the significance of these interactions, which are likely also to occur in vivo (32), remains unclear. By affinity methods, no soluble *E. coli* protein was found to interact with PifA or T7 gp10, but T7 gp1.2 bound *E. coli* dGTPase and PNPase. The former interaction reflects only the inhibition of dGTPase activity by gp1.2 (5, 105, 198) and not F exclusion (188). The significance of the interaction between gp1.2 and PNPase is unclear. Neither polymerization nor RNA degradation by PNPase is affected by gp1.2, and levels of PifA, T7 gp1.2, and gp10 are unaltered by a *pnp* null mutation. However, plating of T7 on *E. coli pnp* mutants containing F is increased by several orders of magnitude, relative to plating on F-plasmid-containing *E. coli* wild-type (32).

References

1. Adhya, S., S. Basu, P. Sarkar, and U. Maitra. 1981. Location, function, and nucleotide sequence of a promoter for bacteriophage T3 RNA polymerase. Proc. Natl. Acad. Sci. USA 78:147–151.
2. Atanasiu, C., O. Byron, H. McMiken, S. S. Sturrock, and D. T. F. Dryden. 2001. Characterisation of the structure of ocr, the gene *0.3* protein of bacteriophage T7. Nucleic Acids Res. 29:3059–3068.
3. Atanasiu, C., T.-J. Su, S. S. Sturrock and D. T. F. Dryden. 2002. Interaction of the ocr gene *0.3* protein of bacteriophage T7 with *Eco*KI restriction/modification enzyme. Nucliec Acids Res. 30:3936–3944.
4. Bandwar, R. P., Y. Jia, N. Stano, and S. S. Patel. 2002. Kinetic and thermodynamic basis of promoter strength: multiple steps of transcription initiation by T7 RNA polymerase are modulated by the promoter sequence. Biochemistry 41:3586–3595.
5. Beauchamp, B. B., and C. C. Richardson. 1988. A unique deoxyguanosinetriphosphatase is responsible for the OptA1 phenotype of *Escherichia coli*. Proc. Natl. Acad. Sci. USA 85:2563–2567.
6. Beck, P. J., J. P. Condreay, and I. J. Molineux. 1986. Expression of the unassembled capsid protein during infection of *Shigella sonnei* by bacteriophage T7 results in DNA damage that is repairable by bacteriophage T3, but not T7, DNA ligase. J. Bacteriol. 167:251–256.
7. Beck, P. J., S. Gonzalez, C. L. Ward, and I. J. Molineux. 1989. Sequence of bacteriophage T3 DNA from gene *2.5* through gene *9*. J. Mol. Biol. 210:687–701.
8. Beck, P. J., and I. J. Molineux. 1991. Defective transcription of the right end of bacteriophage T7 DNA during an abortive infection of F plasmid-containing *Escherichia coli*. J. Bacteriol. 173:947–954.

9. Bedford, E., S. Tabor, and C. C. Richardson. 1997. The thioredoxin binding domain of bacteriophage T7 DNA polymerase confers processivity on *Escherichia coli* DNA polymerase I. Proc. Natl. Acad. Sci. USA 94:479–484.
10. Bernstein, J. A., and C. C. Richardson. 1988. Purification of the 56-kDa component of the bacteriophage T7 primase/helicase and characterization of its nucleoside 5′-triphosphatase activities. J. Biol. Chem. 263:14891–14899.
11. Bernstein, J. A., and C. C. Richardson. 1989. Characterization of the helicase and primase activities of the 63-kDa component of the bacteriophage T7 gene *4* protein. J. Biol. Chem. 264:13066–13073.
12. Blumberg, D. D., C. T. Mabie, and M. H. Malamy. 1976. T7 protein synthesis in F-factor-containing cells: evidence for an episomally induced impairment of translation and its relation to an alteration in membrane permeability. J. Virol. 17:94–105.
13. Blumberg, D. D., and M. H. Malamy. 1974. Evidence for the presence of nontranslated T7 late mRNA in infected F (PIF+) episome-containing cells. J. Virol. 13:378–385.
14. Bonner, G., E. M. Lafer, and R. Sousa. 1994. Characterization of a set of T7 RNA polymerase active site mutants. J. Biol. Chem. 269:25120–25128.
15. Briat, J. F., G. Bollag, C. A. Kearney, I. Molineux, and M. J. Chamberlin. 1987. Tau factor from *Escherichia coli* mediates accurate and efficient termination of transcription at the bacteriophage T3 early termination site in vitro. J. Mol. Biol. 198:43–49.
16. Briat, J.-F., and M. J. Chamberlin. 1984. Identification and characterization of a new transcriptional termination factor from *Escherichia coli*. Proc. Natl. Acad. Sci. USA 81:7373–7377.
17. Bull, J. J., M. R. Badgett, and I. J. Molineux. 2001. A general mechanism for viral resistance to suicide gene expression. J. Mol. Evol. 53:47–54.
18. Bull, J. J., M. R. Badgett, D. Rokyta, D., and I. J. Molineux. 2003. Experimental evolution yields hundreds of mutations in a functional viral genome. J. Mol. Evol. 57:241–248.
19. Bull, J. J., C. Cunningham, I. J. Molineux, M. R. Badgett, and D. M. Hillis. 1993. Experimental molecular evolution using bacteriophage T7. Evolution 47:993–1007.
20. Campbell, J. L., C. C. Richardson, and F. W. Studier. 1978. Genetic recombination and complementation between bacteriophage T7 and cloned fragments of T7 DNA. Proc. Natl. Acad. Sci. USA 75:2276–2280.
21. Casjens, S., K. Eppler, R. Parr, and A. R. Poteete. 1989. Nucleotide sequence of the bacteriophage P22 gene *19* to *3* region: identification of a new gene required for lysis. Virology 171:588–598.
22. Center, M. S., and C. C. Richardson. 1970. An endonuclease induced after infection of *Escherichia coli* with bacteriophage T7. I. Purification and properties of the enzyme. J. Biol. Chem. 245:6285–6291.
23. Cerritelli, M. E., B. Cheng, A. H. Rosenberg, C. E. McPherson, F. P. Booy, and A. C. Steven. 1997. Encapsidated conformation of bacteriophage T7 DNA. Cell 91:271–280.
24. Cerritelli, M . E., and F. W. Studier. 1996. Assembly of T7 capsids from independently expressed and purified head protein and scaffolding protein. J. Mol. Biol. 258:286–298.
25. Cerritelli, M . E., and F. W. Studier. 1996. Purification and characterization of T7 head–tail connectors expressed from the cloned gene. J. Mol. Biol. 258:299–307.
26. Cerritelli, M. E., B. L. Trus, C. S. Smith, N. Cheng, J. F. Conway, and A. C. Steven. 2003. A second symmetry mismatch at the portal vertex of bacteriophage T7: 8-fold symmetry in the procapsid core. J. Mol. Biol. 327:1–6.
27. Chamberlin, M., J. McGrath, and L. Waskell. 1970. New RNA polymerase from *Escherichia coli* infected with bacteriophage T7. Nature 228:227–231.
28. Chamberlin, M. J., W. C. Nierman, J. Wiggs, and N. Neff. 1979. A quantitative assay for bacterial DNA polymerases. J. Biol. Chem. 254:10061–10069.
29. Chang, C.-Y., and I. J. Molineux. Unpublished observations.
30. Cheetham, G. M. T., D. Jeruzalmi, and T. A. Steitz. 1999. Structural basis for initiation of transcription from an RNA polymerase–promoter complex. Nature 399:80–83.
31. Cheetham, G. M. T., and T. A. Steitz. 1999. Structure of a transcribing T7 RNA polymerase initiation complex. Science 286:2305–2309.
32. Cheng, X. 2002. F exclusion of bacteriophage T7. PhD thesis, University of Texas at Austin.
33. Cheng, X., X. Zhang, J. W. Pflugrath, and F. W. Studier. 1994. The structure of bacteriophage T7 lysozyme, a zinc amidase and an inhibitor of T7 RNA polymerase. Proc. Natl. Acad. Sci. USA 91:4034–4038.
34. Christiansen, J. 1988. The 9S RNA precursor of *Escherichia coli* 5S RNA has three structural domains: implication for processing. Nucleic Acids Res. 16:7457–7476.
35. Christie, G. E., and I. J. Molineux. Unpublished observations.
36. Chung, Y.-B., and D. C. Hinkle. 1990. Bacteriophage T7 DNA packaging. I. Plasmids containing a T7 replication origin and the T7 concatemer junction are packaged into transducing particles during phage infection. J. Mol. Biol. 216:911–926.
37. Chung, Y.-B., and D. C. Hinkle. 1990. Bacteriophage T7 DNA packaging. II. Analysis of the DNA sequences required for packaging using a plasmid transduction assay. J. Mol. Biol. 216:927–938.
38. Chung, Y.-B., C. Nardone, and D. C. Hinkle. 1990. Bacteriophage T7 DNA packaging. III A "hairpin" end formed on T7 DNA concatemers may be an intermediate in the processing reaction. J. Mol. Biol. 216:939–948.
39. Clarke, B. R., F. Esumeh, and I. S. Roberts. 2000. Cloning, expression, and purification of the K5 capsular polysaccharide lyase (KflA) from coliphage K5A: evidence for two distinct K5 lyase enzymes. J. Bacteriol. 182:3761–3766.
40. Condit, R. C., and J. A. Steitz. 1975. F factor-mediated inhibition of bacteriophage T7 growth: analysis of T7 RNA and protein synthesis in vivo and in vitro using male and female *Escherichia coli*. J. Mol. Biol. 98:31–43.
41. Condit, R. C. 1975. F factor-mediated inhibition of bacteriophage T7 growth: increased membrane permeability and decreased ATP levels following T7 infection of male *Escherichia coli*. J. Mol. Biol. 98:45–56.
42. Condreay, J. P., and I. J. Molineux. 1989. Synthesis of the capsid protein inhibits development of bacteriophage T3 mutants that abortively infect F plasmid-containing strains. J. Mol. Biol. 207:543–554.

43. Condreay, J. P., S. E. Wright, and I. J. Molineux. 1989. Nucleotide sequence and complementation studies of the gene *10* region of bacteriophage T3. J. Mol. Biol. 207:555–561.
44. Condron, B. G., J. F. Atkins, and R. F. Gesteland RF. 1991. Frameshifting in gene *10* of bacteriophage T7. J. Bacteriol. 173:6998–7003.
45. Condron, B. G., R. F. Gesteland, and J. F. Atkins. 1991. An analysis of sequences stimulating frameshifting in the decoding of gene *10* of bacteriophage T7. Nucleic Acids Res. 19:5607–5612.
46. Cram, H. K., D. Cram, and R. Skurray. 1984. F plasmid *pif* region: Tn*1725* mutagenesis and polypeptide analysis. Gene 32:251–254.
47. Cram, D., A. Ray, and R. Skurray. 1984. Molecular analysis of F plasmid *pif* region specifying abortive infection of T7. Mol. Gen. Genet. 197:137–142.
48. Cunningham, C. W., K. Jeng, J. Husti, M. Badgett, I. J. Molineux, D. M. Hillis, and J. J. Bull. 1997. Parallel molecular evolution of deletions and nonsense mutations in bacteriophage T7. J. Mol. Evol. 14:113–116.
49. Dausse, J.-P., A. Sentenac, and P. Fromageot. 1976. Interaction of RNA polymerase from *Escherichia coli* with DNA. Eur. J. Biochem. 65:387–393.
50. Déclais, A., J. M. Fogg, A. D. J. Freeman, F. Coste, J. M. Hadden, S. E. V. Phillips, and D. M. J. Lilley. 2003. The complex between a four-way DNA junction and T7 endonuclease I. EMBO J. 22:1398–1409.
51. De Massy, B., R. A. Weisberg, and F. W. Studier. 1987. Gene *3* endonuclease of bacteriophage T7 resolves conformationally branched structures in double-stranded DNA. J. Mol. Biol. 193:359–376.
52. Demchick, P., and A. L. Koch. 1996. The permeability of the wall fabric of *Escherichia coli* and *Bacillus subtilis.* J. Bacteriol. 178:768–773.
53. Demerec, M., and U. Fano. 1945. Bacteriophage-resistant mutants in *Escherichia coli.* Genetics 30:119–136.
54. De Wyngaert, M. A., and D. C. Hinkle. 1979. Bacterial mutants affecting phage T7 replication produce RNA polymerase resistant to inhibition by the T7 gene *2* protein. J. Biol. Chem. 254:11247–11253.
55. De Wyngaert, M. A., and D. C. Hinkle. 1980. Characterization of the defects in bacteriophage T7 DNA synthesis during growth in the *Escherichia coli* mutant *tsnB*. J. Virol. 33:780–788.
56. Dobbins, A. T., M. George Jr., D. J. Basham, M. E. Ford, J. M. Houtz, M. L. Pedulla, J. G. Lawrence, G. F. Hatfull, and R. W. Hendrix. 2004. Complete genomic sequence of the virulent *Salmonella* bacteriophage SP6. J. Bacteriol. 186:1933–1944.
57. Donate, L. E., L. Herranz, J. P. Secilla, J. M. Carazo, H. Fujisawa, and J. L. Carrascosa. 1988. Bacteriophage T3 connector: three-dimensional structure and comparison with other viral head–tail connecting regions. J. Mol. Biol. 201:91–100.
58. Doublié, S., S. Tabor, A. M. Long, C. C. Richardson, and T. Ellenberger. 1998. Crystal structure of a bacteriophage T7 DNA replication complex at 2.2Å resolution. Nature 391:251–258.
59. Dressler, D. J. Wolfson, and M. Magazin. 1972. Initiation and reinitiation of DNA synthesis during replication of bacteriophage T7. Proc. Natl. Acad. Sci. USA 69:998–1002.
60. Dunn, J. J., and F. W. Studier. 1973. T7 early RNAs and *Esherichia coli* ribosomal RNAs are cut from large precursor RNAs in vivo by ribonuclease III. Proc. Natl. Acad. Sci. USA 70:3296–3300.
61. Dunn, J. J., and F. W. Studier. 1975. Effect of RNase III cleavage on translation of bacteriophage T7 messenger RNAs. J. Mol. Biol. 99:487–499.
62. Dunn, J. J., and F. W. Studier. 1983. Complete nucleotide sequence of bacteriophage T7 DNA and the locations of genetic elements. J. Mol. Biol. 166:477–535.
63. Ederth, J., L. A. Isaksson, and F. Abdulkarim. 2002. Origin-specific reduction of ColE1 plasmid copy number due to mutations in a distinct region of the *Escherichia coli* RNA polymerase. Mol. Genet. Genomics 267:587–592.
64. Egelman, H. H., X. Yu, R. Wild, M. M. Hingorami, and S. S. Patel. 1995. Bacteriophage T7 DNA helicase/primase proteins form rings around single-stranded DNA that suggests a structure for hexameric helicases. Proc. Natl. Acad. Sci. USA 92:3869–3873.
65. Endy, D., L. You, J. Yin, and I. J. Molineux. 2000. Computation, prediction, and experimental tests of fitness for bacteriophage T7 mutants with permuted genomes. Proc. Natl. Acad. Sci USA 97:5375–5380.
66. Feng C., and J. Lu. 2002. Genomic sequence and evolution of marine cyanophage P60: a new insight on lytic and lysogenic phages. Appl. Env. Microbiol. 68:2589–2594.
67. Frick, D. N., K. Baradaran, and C. C. Richardson. 1998. An N-terminal fragment of the gene *4* helicase/primase of bacteriophage T7 retains primase activity in the absence of helicase activity. Proc. Natl. Acad. Sci USA 95:7957–7962.
68. Frick, D. N., and C. C. Richardson. 1999. Interaction of the bacteriophage T7 gene *4* primase with its template recognition site. J. Biol. Chem. 274:35889–35898.
69. Frick, D. N., and C. C. Richardson. 2001. DNA primases. Annu Rev. Biochem. 70:39–80.
70. Frick, D. N., S. Kumar, and C. C. Richardson. 1999. Interaction of ribonucleoside triphosphates with the gene *4* primase of bacteriophage T7. J. Biol. Chem. 274:35899–35907.
71. Fujisawa, H., M. Kimura, and C. Hashimoto. 1990. In vitro cleavage of the concatemer joint of bacteriophage T3. Virology 174:26–34.
72. Fujisawa, H., J. Miyazaki, and T. Minagawa. 1978. In vitro packaging of phage T3 DNA. Virology 87:394–400.
73. Fujisawa, H., and M. Morita. 1997. Phage DNA packaging. Genes Cells 2:537–545.
74. Fujisawa, H., H. Shibata, and H. Kato. 1991. Analysis of interactions among factors involved in the bacteriophage T3 DNA packaging reaction in a defined in vitro system. Virology 185:788–794.
75. Fujisawa, H., M. Yamagishi, and T. Minagawa. 1980. In vitro formation of the concatemeric DNA of bacteriophage T3 and its biological activity in the in vitro packaging reaction. Virology 101:327–334.

76. Fuller, C. W., and C. C. Richardson. 1985. Initiation of DNA replication at the primary origin of bacteriophage T7 by purified proteins. Site and direction of initial DNA synthesis. J. Biol. Chem. 260:3185–3196.
77. Fuller, C. W., and C. C. Richardson. 1985. Initiation of DNA replication at the primary origin of bacteriophage T7 by purified proteins. Initiation of bidirectional synthesis. J. Biol. Chem. 260:3197–3206.
78. Garcia, E., J. M. Elliott, E. Ramanculov, P. S. G. Chain, M. C. Chu, and I. J. Molineux. The genome sequence of *Yersinia pestis* bacteriophage ϕA1122 reveals an intimate history with the coliphage T3 and T7 genomes. J. Bacteriol. 185:5248–5262.
79. García, L. R., and I. J. Molineux. 1995. Rate of translocation of bacteriophage T7 DNA across the membranes of *Escherichia coli.* J. Bacteriol. 177:4066–4076.
80. García, L. R., and I. J. Molineux. 1995. Incomplete entry of bacteriophage T7 DNA into F plasmid–containing *Escherichia coli* strains. J. Bacteriol. 177:4077–4083.
81. García, L. R., and Molineux, I. J. 1996. Transcription-independent DNA translocation of bacteriophage T7 DNA into *Escherichia coli.* J. Bacteriol. 178:6921–6929.
82. García, L. R., and Molineux, I. J. 1999. Translocation and cleavage of bacteriophage T7 DNA by the type I restriction enzyme *Eco*KI in vivo. Proc. Natl. Acad. Sci USA 96:12430–12435.
83. Gold, L., D. Pribnow, T. Schneider, S. Shinedling, B. S. Singer, and G. Stormo. 1981. Translational initiation in prokaryotes. Annu. Rev. Microbiol. 35:365–403.
84. Gomez, B., and L. Nualart. 1977. Requirement of the bacteriophage T7 *0.7* gene for phage growth in the presence of the Col Ib factor. J. Gen. Virol. 35:99–106.
85. Grippo, P., and C. C. Richardson. 1971. Deoxyribonucleic acid polymerase of bacteriophage T7. J. Biol. Chem. 246: 6867–6873.
86. Hadden, J. M., M. A. Convery, A.-C. Déclais, D. M. J. Lilley, and S. E. V. Phillips. 2001. Crystal structure of the Holliday junction resolving enzyme T7 endonuclease I. Nature Struct. Biol. 8:62–67.
87. Haggård-Ljungquist, E., Halling, C., and Calendar, R. 1992. DNA sequences of the tail fiber genes of bacteriophage P2: evidence for horizontal transfer of tail fiber genes among unrelated bacteriophages. J. Bacteriol. 174: 1462–1477.
88. Hamada, K., H. Fujisawa, and T. Minagawa. 1986. A defined in vitro system for packaging of bacteriophage T3 DNA. Virology 151:119–123.
89. Hamada, K., H. Fujisawa, and T. Minagawa. 1986. Overproduction and purification of the products of T3 genes *18* and *19*, two genes involved in DNA packaging. Virology 151:110–118.
90. Hardies, S. C., A. M. Comeau, P. Serwer, and C. A. Suttle. 2003. The complete sequence of marine bacteriophage VpV262 infecting *Vibrio parahaemolyticus* indicates that an ancestral component of a T7 viral supergroup is widespread in the marine environment. Virology 310:359–371.
91. Hartvig, L., and J. Christiansen. 2000. Intrinsic termination of T7 RNA polymerase mediated by either RNA or DNA. EMBO J. 15:4767–4774.
92. Hashimoto, C., and H. Fujisawa. 1992. DNA sequences necessary for packaging bacteriophage T3 DNA. Virology 187:788–795.
93. Hashimoto, C., and H. Fujisawa. 1992. Transcription dependence of DNA packaging of bacteriophages T3 and T7. Virology 191:246–250.
94. Hausmann, R. 1968. Sedimentation analysis of phage T7-directed DNA synthesized in the presence of a dominant lethal gene. Biochem. Biophys. Res. Commun. 31:609–615.
95. Hausmann, R. 1988. The T7 group, pp. 259–289. *In* R. Calender (ed.) The Bacteriophages, vol. 1. Plenum Press, New York.
96. Hausmann, R., and B. Gomez. 1968. Bacteriophage T3- and T7-directed deoxyribonucleases. J. Virol. 2:265–266.
97. Hesselbach, B. A., and D. Nakada. 1975. Inactive complex formation between *E. coli* RNA polymerase and inhibitor protein purified from T7 phage infected cells. Nature 258:354–357.
98. Hesselbach, B. A., and D. Nakada. 1977. "Host shut off" function of bacteriophage T7: involvement of T7 gene *2* and gene *0.7* in the inactivation of *Escherichia coli* RNA polymerase. J. Virol. 24:736–745.
99. Hesselbach, B. A., and D. Nakada. 1977. I protein: bacteriophage-coded inhibitor of *Escherichia coli* RNA polymerase. J. Virol. 24:746–760.
100. Hillis, D. M., J. J. Bull, M. E. White, M. R. Badgett, and I. J. Molineux. 1992. Experimental phylogenetics: generation of a known phylogeny. Science 255:589–592.
101. Hingorani, M. M., M. T. Washington, K. C. Moore, and S. S. Patel. 1997. The dTTPase mechanism of T7 DNA helicase resembles the binding change mechanism of the F_1-ATPase. Proc. Natl. Acad. Sci. USA 94:5012–5017.
102. Hirsch-Kauffmann, M., P. Herrlich, H. Ponta, and M. Schweiger. 1975. Helper function of T7 protein kinase in virus propagation. Nature 255:508–510.
103. Hollis, T., J. M. Stattel, D. S. Walther, C. C. Richardson, and T. Ellenberger. 2001. Structure of the gene *2.5* protein, a single-stranded DNA binding protein encoded by bacteriophage T7. Proc. Natl. Acad. Sci. USA 98: 9557–8562.
104. Huang, J., J. Villemain, R. Padilla, and R. Sousa. 1999. Mechanisms by which T7 lysozyme specifically regulates T7 RNA polymerase during different phases of transcription. J. Mol. Biol. 293:457–475.
105. Huber, H. E., B. B. Beauchamp, and C. C. Richardson. 1988. *Escherichia coli* dGTP triphosphohydrolase is inhibited by gene *1.2* protein of bacteriophage T7. J. Biol. Chem. 263:13549–13556.
106. Huber, H. E., M. Russel, P. Model, and C. C. Richardson. 1986. Interaction of mutant thioredoxins of *Escherichia coli* with the gene *5* protein of phage T7. The redox capacity of thioredoxin is not required for stimulation of DNA polymerase activity. J. Biol. Chem. 261:15006–15012.
107. Huber, H. E., S. Tabor, and C. C. Richardson. 1987. *Escherichia coli* thioredoxin stabilizes complexes of bacteriophage T7 DNA polymerase and primed templates. J. Biol. Chem. 262:16224–16232.

108. Hughes, V., and G. G. Meynell. 1973. The contribution of plasmid and phage genes to plasmid-mediated interference with phage growth. Genet. Res. 30:179–185.
109. Ikeda, R. A., and P. A. Bailey. 1992. Inhibition of T7 RNA polymerase by T7 lysozyme in vitro. J. Biol. Chem. 267:20153–20158.
110. Ikeda, R. A., and C. C. Richardson. 1986. Interactions of the RNA polymerase of bacteriophage T7 with its promoter during binding and initiation of transcription. Proc. Natl. Acad. Sci. USA 83:3614–3618.
111. Imburgio, D., M. Rong, K. Ma, and W. T. McAllister. 2000. Studies of promoter recognition and start site selection by T7 RNA polymerase using a comprehensive collection of promoter variants. Biochemistry 39:10419–10430.
112. Inouye, M., N. Arnheim, and R. Sternglanz. 1973. Bacteriophage T7 lysozyme is an N-acetylmuramy-L-alanine amidase. J. Biol. Chem. 248:7247–7252.
113. Jeng, S. T., J. F. Gardner, and R. I. Gumport. 1990. Transcription termination by bacteriophage T7 RNA polymerase at rho-independent terminators. J. Biol. Chem. 265: 3823–3830.
114. Jeruzalmi, D., and T. A. Steitz. 1998. Structure of T7 RNA polymerase complexed to the transcriptional inhibitor T7 lysozyme. EMBO J. 14:4101–4113.
115. Kato, H., H. Fujisawa, and T. Minagawa. 1986. Subunit arrangement of the tail fiber of bacteriophage T3. Virology 153:80–86.
116. Kato, M., T. Ito, G. Wagner, C. C. Richardson, and T. Ellenberger. 2003. Modular architecture of the bacteriophage T7 primase couples RNA primer synthesis to DNA synthesis. Mol. Cell 11:1349–1360.
117. Kelly, T. J., Jr., and C. A. Thomas, Jr. 1969. An intermediate in the replication of bacteriophage T7 DNA molecules. J. Mol. Biol. 44:459–475.
118. Kemp, P., M. Gupta, and I. J. Molineux. 2004. Bacteriophage T7 DNA ejection into cells is initiated by an enzyme-like process. Mol. Microbiol. 53:1251–1265.
119. Kemp, P., and I. J. Molineux. Unpublished observations.
120. Kemp, P., Chang, C.-Y., L. R. García, and I. J. Molineux. Unpublished observations.
121. Kennedy, M., M. Chandler, and D. Lane. 1988. Mapping and regulation of the *pifC* promoter of the F plasmid. Biochim. Biophys. Acta 950:75–80.
122. Kerr, C., and P. D. Sadowski. 1975. The involvement of genes *3, 4, 5* and *6* in genetic recombination of bacteriophage T7. Virology 65:281–285.
123. Kiefer, M., N. Neff, and M. J. Chamberlin. 1977. Transcriptional termination at the end of the early region of bacteriophages T3 and T7 is not affected by polarity suppressors. J. Virol. 22:548–552.
124. Kieleczawa, J., and I. J. Molineux. Unpublished observation.
125. Kim, D.-E., M. Narayan, and S. S. Patel. 2002. T7 DNA helicase: a molecular motor that processively and unidirectionally translocates along single-stranded DNA. J. Mol. Biol. 321:807–819.
126. Kim, S.-H., and Y.-B. Chung. 1996. Isolation of a mutant bacteriophage T7 deleted in nonessential genetic elements, gene *19.5* and *m*. Virology 216:20–25.
127. Kim, J.-S., S.-H. Kim, and Y.-B. Chung. 1997. Defects in concatemer processing of bacteriophage T7 DNA deleted in the m-hairpin region. Virology 236:37–46.
128. Kim, Y. T., and C. C. Richardson. 1993. Bacteriophage T7 gene *2.5* protein: an essential protein for DNA replication. Proc. Natl. Acad. Sci. USA 90:10173–10177.
129. Kim, Y. T., and C. C. Richardson. 1994. Acidic carboxyl-terminal domain of gene *2.5* protein of bacteriophage T7 is essential for protein–protein interactions. J. Biol. Chem. 269:5270–5278.
130. Kim, Y. T., S. Tabor, J. E. Churchich, and C. C. Richardson. 1992. Interactions of gene *2.5* protein and DNA polymerase of bacteriophage T7. J. Biol. Chem. 267: 15032–15040.
131. Klement, J. F., M. B. Moorefield, E. Jorgensen, J. E. Brown, S. Risman, and W. T. McAllister. 1990. Discrimination between bacteriophage T3 and T7 promoters by the T3 and T7 RNA polymerases depends primarily upon a three base-pair region located 10 to 12 base-pairs upstream from the start site. J. Mol. Biol. 215:21–29.
132. Kocsis, E., M. E. Cerritelli, B. L. Trus, N. Cheng, and A. C. Steven. 1995. Improved methods for determination of rotational symmetries in macromolecules. Ultramicroscopy 60:219–228.
133. Kong, D., N. G. Nossal, and C. C. Richardson. 1997. Role of the bacteriophage T7 and T4 single-stranded DNA-binding proteins in the formation of joint molecules and DNA helicase-catalyzed polar branch migration. J. Biol. Chem. 272:8380–8387.
134. Kong, D., and C. C. Richardson. 1996. Single-stranded DNA binding protein and DNA helicase of bacteriophage T7 mediate homologous DNA strand exchange. EMBO J. 15:2010–2019.
135. Kovalayova, I. V., and A. M. Kropinski. 2003. The complete genome sequence of lytic bacteriophage gh-1 infecting *Pseudomonas putida*: evidence for close relationship to the T7 group. Virology 311:305–315.
136. Kreuzer, K. N., and N. R. Cozzarelli. 1979. *Escherichia coli* mutants themosensitive for deoxyribonucleic acid gyrase subunit A: effects on deoxyribonucleic acid replication, transcription and bacteriophage growth. J. Bacteriol. 140:424–435.
137. Krüger, D. H., and C. Schroeder. 1981. Bacteriophage T3 and bacteriophage T7 virus–host cell interactions. Microbiol. Rev. 45:9–51.
138. Kuhn, A., and E. Kellenberger. 1985. Productive phage infection in *Escherichia coli* with reduced internal levels of the major cations. J. Bacteriol. 163:906–912.
139. Kumar, A., and S. S. Patel. 1997. Inhibition of T7 RNA polymerase transcription initiation and transition from initiation to elongation are inhibited by T7 lysozyme via a ternary complex with RNA polymerase and promoter DNA. Biochemistry 36:13954–13962.
140. Kwiatkowski, B., B. Boschek, H. Thiele, and S. Stirm. 1982. Endo-*N*-acetylneuraminidase associated with bacteriophage particles. J. Virol. 43:697–704.
141. Kwiatkowski, B., B. Boschek, H. Thiele H, and S. Stirm. 1983. Substrate specificity of two bacteriophage-associated endo-*N*-acetylneuraminidases. J. Virol. 45:367–374.

142. Langman, L., and V. Paetkau. 1978. Purification and structures of recombining and replicating bacteriophage T7 DNA. J. Virol. 25:562–569.
143. Langman, L., V. Paetkau, D. Scraba, R. C. Miller, Jr., G. S. Roeder, and P. D. Sadowski. 1978. The structure and maturation of intermediates in bacteriophage T7 DNA replication. Can. J. Biochem. 56:508–516.
144. Lavigne, R., M. V. Bourkaltseva, J. Robben, N. N.. Sykilinda, L. P. Kurochkina, B. Grymonprez, B. Jonckx, V. N. Krylov, V. V. Mesyanzhinov, and G. Volckaert. 2003. The genome of bacteriophage ϕKMV: a T7-like lytic phage infecting *Pseudomonas aeruginosa*. Virology 312:49–59.
145. Lazarus, A. S., and J. B. Gunnison. 1947. The action of *Pasteurella pestis* bacteriophage on strains of *Pasteurella*, *Salmonella*, and *Shigella*. J. Bacteriol. 53:705–714.
146. LeClerc, J. E., and C. C. Richardson. 1979. Bacteriophage T7 DNA replication in vitro. XVI. Gene 2 protein of bacteriophage T7: purification and requirement for packaging of T7 DNA in vitro. Proc. Natl. Acad. Sci. USA 76:4852–4856.
147. Lee, D., and P. D. Sadowski. 1981. Genetic recombination of bacteriophage T7 in vivo studied by use of a simple physical assay. J. Virol. 40:839–847.
148. Lee, J., P. D. Chastain II, J. D. Griffith, and C. C. Richardson. 2002. Lagging strand synthesis in coordinated DNA synthesis by T7 replication proteins. J. Mol. Biol. 316:19–34.
149. Lee, J., P. D. Chastain II, T. Kusakabe, J. D. Griffith, and C. C. Richardson. 1998. Coordinated leading and lagging strand synthesis on a mini-circular template. Mol. Cell 1:1001–1010.
150. Lee, M., and R. C. Miller, Jr. 1974. T7 exonuclease (gene *6*) is necessary for molecular recombination of bacteriophage T7. J. Virol. 14:1040–1048.
151. Lee, M., R. C. Miller, Jr., D. Scraba, and V. Paetkau. 1976. The essential role of bacteriophage endonuclease (gene *3*) in molecular recombination. J. Mol. Biol. 104:883–888.
152. Lee, S. S., and K. C. Kang. 1992. A two-base-pair substitution in T7 promoter by SP6 promoter-specific base pairs alone abolishes T7 promoter activity but reveals SP6 promoter activity. Biochem. Int. 26:1–5.
153. Lin, L. 1992. Study of bacteriophage T7 gene *5.9* and gene *5.5*. PhD thesis, SUNY, Stonybrook, New York.
154. Liu, Q., and C. C. Richardson. 1993. Gene 5.5 protein of bacteriophage T7 inhibits the nucleoid protein H-NS of *Escherichia coli*. Proc. Natl. Acad. Sci. USA 90:1761–1765.
155. Long, G. S., J. M. Bryant, P. W. Taylor, and J. P. Luzio. 1995. Complete nucleotide sequence of the gene encoding bacteriophage E endosialidase: implications for K1E endosialidase structure and function. Biochem. J. 309:543–550.
156. Lopez, P. J., J. Guillerez, R. Sousa, and M. Dreyfus. The low processivity of T7 RNA polymerase over the initially transcribed region can limit productive initiation in vivo. J. Mol. Biol. 269:41–51.
157. Lu, M., Q. Guo, F. W. Studier, and N. R. Kallenbach. 1991. Resolution of branched DNA substrates by T7 endonuclease I and its inhibition. J. Biol. Chem. 266:2531–2536.
158. Lyakhov, D. L., B. He, X. Zhang, F. W. Studier, J. J. Dunn, and W. T. McAllister. 1997. Mutant bacteriophage T7 RNA polymerases with altered termination properties. J. Mol. Biol. 269:28–40.
159. Lyakhov, D. L., B. He, X. Zhang, F. W. Studier, J. J. Dunn, and W. T. McAllister. 1998. Pausing and termination by bacteriophage T7 RNA polymerase. J. Mol. Biol. 280: 201–213.
160. Ma, K., D. Temiakov, M. Jiang, M. Anikin, and W. T. McAllister. 2002. Major conformational changes occur during the transition from an initiation complex to an elongation complex by T7 RNA polymerase. J. Biol. Chem. 277:43206–43215.
161. MacDonald, L. E., Y. Zhou, and W. T. McAllister. 1993. Termination and slippage by T7 RNA polymerase. J. Mol. Biol. 232:1030–1047.
162. Macdonald, L. E., R. K. Durbin, J. J. Dunn, and W. T. McAllister. 1994. Characterization of two types of termination signal for bacteriophage T7 RNA polymerase. J. Mol. Biol. 238:145–158.
163. Makarova, O. V., E. M. Makarov, R. Sousa, and M. Dreyfus. 1995. Transcribing of *Escherichia coli* genes with mutant T7 RNA polymerases: stability of *lacZ* mRNA inversely correlates with polymerase speed. Proc. Natl. Acad. Sci. USA 92:12250–12254.
164. Mäkelä, O., P. H. Mäkelä, and S. Soikkeli. 1964. Sex-specificity of the bacteriophage T7. Ann. Med. Exp. Biol. Fenn. 42:188–195.
165. Marchand, I., A. W. Nicholson, and M. Dreyfus. 2001. Bacteriophage T7 protein kinase phosphorylates RNase E and stabilizes mRNAs synthesized by T7 RNA polymerase. Mol. Microbiol. 42:767–776.
166. Marrs, B. L., and C. Yanofsky. 1971. Host and bacteriophage specific messenger RNA degradation in T7-infected *Escherichia coli*. Nature New Biol. 234:168–170.
167. Martin, C. T., and J. E. Coleman. 1987. Kinetic analysis of T7 RNA polymerase–promoter interactions with small synthetic promoters. Biochemistry 27:2690–2696.
168. Martin, C. T., D. K. Muller, and J. E. Coleman. 1988. Processivity in early stages of transcription by T7 RNA polymerase. Biochemistry 27:3966–3974.
169. Matson, S. W., and C. C. Richardson. 1983. DNA-dependent nucleoside 5′-triphosphatase activity of the gene *4* protein of bacteriophage T7. J. Biol. Chem. 258:14009–14016.
170. Matson, S. W., S. Tabor, and C. C. Richardson. 1983. The gene *4* protein of bacteriophage T7. Characterization of helicase activity. J. Biol. Chem. 258:14017–14024.
171. Matsuo-Kato, H., H. Fujisawa, and T. Minagawa. 1981. Structure and assembly of bacteriophage T3 tails. Virology 109:157–164.
172. Mayer, J. E., and M. Schweiger. 1983. RNase III is positively regulated by T7 protein kinase. J. Biol. Chem. 258:5340–5343.
173. McAllister, W. T., and C. L. Barrett. 1977. Roles of the early genes of bacteriophage T7 in shutoff of host macromolecular synthesis. J. Virol. 28:543–553.
174. McAllister, W. T., C. Morris, A. H. Rosenberg, and F. W. Studier. 1981. Utilization of bacteriophage T7 late

promoters in recombinant plasmids during infection. J. Mol. Biol. 153:527–544.

175. McAllister, W. T., and H.-L. Wu. 1978. Regulation of transcription of bacteriophage T7. Proc. Natl. Acad. Sci. USA 75:804–808.
176. Meisel, A., T. A. Bickle, D. H. Kruger, and C. Schroeder. 1992. Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. Nature 355:467–469.
177. Mendelman, L. V., S. M. Notarnicola, and C. C. Richardson. 1992. Roles of bacteriophage T7 gene *4* proteins in providing priimase and helicase functions in vivo. Proc. Natl. Acad. Sci. USA 89:10638–10642.
178. Mendelman, L. V., and C. C. Richardson. 1991. Requirements for primer synthesis by bacteriophage T7 63-kDa gene *4* protein. Roles of template sequence and T7 56-kDa gene *4* protein. J. Biol. Chem. 266:23240–23250.
179. Mentesana, P. E., S. T. Chin-Bow, R. Sousa, and W. T. McAllister. 2000. Characterization of halted RNA polymerase elongation complexes reveals multiple factors that contribute to stability. J. Mol. Biol. 302:1049–1062.
180. Meynell, E., G. G. Meynell, and N. Datta. 1968. Phylogenetic relationships of drug-resistance factors and other transmissible plasmids. Bacteriol. Rev. 32:55–83.
181. Miller, J. F., and M. H. Malamy. 1983. Identification of the *pifC* gene and its role in negative control of F factor gene expression. J. Bacteriol. 156:338–347.
182. Miller, J. F., and M. H. Malamy. 1984. Regulation of the F-factor *pif* operon: *pifO*, a site required in *cis* for autoregulation, titrates the *pifC* product in *trans*. J. Bacteriol. 160:192–198.
183. Moak, M., and I. J. Molineux. 2000. Role of the gp16 lytic tranglycosylase motif in bacteriophage T7 virions at the initiation of infection. Mol. Microbiol. 37:345–355.
184. Moak, M., and I. J. Molineux. 2004. Peptidoglycan hydrolytic activities associated with bacteriophage virions. Mol. Microbiol. 51:1169–1183.
185. Moffat, B. A., J. J. Dunn, and F. W. Studier. 1984. Nucleotide sequence of the gene for bacteriophage T7 RNA polymerase. J. Mol. Biol. 173:265–269.
186. Moffatt, B. A., and F. W. Studier. 1987. T7 lysozyme inhibits transcription by T7 RNA polymerase. Cell 49:221–227.
187. Moffatt, B. A., and F. W. Studier. 1988. Entry of bacteriophage T7 DNA into the cell and escape from host restriction. J. Bacteriol. 170:2095–2105.
188. Molineux, I. J. 1991. Host–parasite interactions: recent developments in the genetics of abortive phage infections. New Biol. 3:230–236.
189. Molineux, I. J. 2001. No syringes please, ejection of T7 DNA from the virion is enzyme-driven. Mol. Microbiol. 40:1–8.
190. Molineux, I. J., C. K. Schmitt, and J. P. Condreay. 1989. Mutants of bacteriophage T7 that escape F restriction. J. Mol. Biol. 207:563–574.
191. Molineux, I. J., and J. L. Spence. 1984. Virus–plasmid interactions: mutants of bacteriophage T3 that abortively infect plasmid F-containing (F^+) strains of *Escherichia coli*. Proc. Natl. Acad. Sci. USA 81:1465–1469.
192. Mooney, P. Q., North, R., and I. J. Molineux. 1980. The role of bacteriophage T7 gene *2* protein in DNA replication. Nucleic Acids Res. 8:3043–3053.
193. Morita, M., M. Tasaka, and H. Fujisawa. 1993. DNA packaging ATPase of bacteriophage T3. Virology 193:748–752.
194. Morita, M., M. Tasaka, and H. Fujisawa. 1994. Analysis of functional domains of the packaging proteins of bacteriophage T3 by site-directed mutagenesis. J. Mol. Biol. 235:248–259.
195. Morita, M., M. Tasaka, and H. Fujisawa. 1995. Structural and functional domains of the DNA packaging protein of bacteriophage T3: importance of the C-terminal region of the large subunit in prohead binding. J. Mol. Biol. 245:635–644.
196. Morrison, T. G., D. D. Blumberg, and M. H. Malamy. 1974. T7 protein synthesis in F′ episome-containing cells: assignment of specific proteins to three translational groups. J. Virol. 13:386–393.
197. Morrison, T. G., and M. H. Malamy. 1971. T7 translational control mechanisms and their inhibition by F factors. Nature New Biol. 231:37–41.
198. Nakai, H., and C. C. Richardson. 1990. The gene *1.2* protein of bacteriophage T7 interacts with the *Escherichia coli* dGTP triphosphohydrolase to form a GTP-binding protein. J. Biol. Chem. 265:4411–4419.
199. Nakasu, S., H. Fujisawa, and T. Minagawa. 1983. Role of gene *8* product in morphogenesis of bacteriophage T3. Virology 127:124–133.
200. Nechaev, S., and K. Severinov. 1999. Inhibition of *Escherichia coli* RNA polymerase by bacteriophage T7 gene *2* protein. J. Mol. Biol. 289:815–826.
201. Nelson, K., et al. 2002. Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. Environ. Microbiol. 4:799–808.
202. Noji, H., and M. Yoshida. 2001. The rotary machine in the cell, ATP synthase. J. Biol. Chem. 276:1665–1668.
203. Notarnicola, S. M., K. Park, J. D. Griffith, and C. C. Richardson. 1995. A domain of the gene *4* helicase/primase of bacteriophage T7 required for the formation of an active hexamer. J. Biol. Chem. 270:20215–20224.
204. Ontell, M. P., and D. Nakada. 1980. Rescue of abortive T7 gene *2* mutant phage infection by rifampin. J. Virol. 34:438–445.
205. Paetkau, V., L. Langman, R. Bradley, D. Scraba, and R. C. Miller. 1977. Folded, concatenated genomes as replication intermediates of bacteriophage T7 DNA. J. Virol. 22:130–141.
206. Pajunen, M. I., M. R. Elizondo, M. Skurnik, J. Kieleczawa, and I. J. Molineux. 2002. Complete nucleotide sequence and likely recombinatorial origin of bacteriophage T3. J. Mol. Biol. 319:1115–1132.
207. Pajunen, M. I., S. J. Kiljunen, M. E.-L. Söderholm, and M. Skurnik. 2001. Complete genomic sequence of the lytic bacteriophage ϕYeO3-12 of *Yersinia enterocolitica* serotype O:3 J. Bacteriol. 183:1928–1937.
208. Pajunen, M. I., I. J. Molineux, and M. Skurnik. Unpublished observation.
209. Patel, S. S., Rosenberg, A. H., K. Griffin, F. W. Studier, and K. A. Johnson. 1992. Large scale purification and

biochemical characterization of T7 primase/helicase proteins. Evidence for homodimers and heterodimers formation. J. Biol. Chem. 267:15013–15021.

210. Petter, J. G., and E. R. Vimr. 1993. Complete nucleotide sequence of the bacteriophage K1F tail gene encoding endo-*N*-acylneuraminidase (Endo-*N*) and comparison to an Endo-*N* homolog in bacteriophage PK1E. J. Bacteriol. 175:4354–4363.

211. Pfennig-Yeh, M. L., H. Ponta, M. Hirsch-Kauffmann, H. J. Rahmsdorf, P. Herrlich, and M. Schweiger. 1978. Early T7 gene expression: rates of RNA synthesis and degradation, protein kinase dependent termination of transcription, and efficiency of translation. Mol. Gen. Genet. 166:127–140.

212. Pierce, J. C., and W. E. Masker. 1988. A single base change in gene *10* of bacteriophage T7 permits growth on *Shigella sonnei*. J. Virol. 62:4369–4371.

213. Prehm, P., B. Jann, K. Jann, G. Schmidt, and S. Stirm. 1976. On a bacteriophage T3 and T4 receptor region within the cell wall lipopolysaccharide of *Escherichia coli* B. J. Mol. Biol. 101:277–281.

214. Rabkin, S. D. and C. C. Richardson. 1988. Initiation of DNA replication at cloned origins of bacteriophage T7. J. Mol. Biol. 204:903–916.

215. Rabkin, S. D. and C. C. Richardson. 1990. In vivo analysis of the initiation of bacteriophage T7 DNA replication. Virology 174:585–592.

216. Rahmsdorf, H. J., S. H. Pai, H. Ponta, P. Herrlich, R. Roskowski, M. Schweiger, and F. W. Studier. 1974. Protein kinase induction in *E. coli* bacteriophage T7. Proc. Natl. Acad. Sci. USA 71:586–589.

217. Ramanculov, E., M. C. Chu, and I. J. Molineux. Unpublished observations.

218. Raskin C. A, G. A. Diaz, K. Joho, and W. T. McAllister. 1992. Substitution of a single bacteriophage T3 residue in bacteriophage T7 RNA polymerase at position 748 results in a switch in promoter specificity. J. Mol. Biol. 228:506–515.

219. Raskin, C. A., G. A. Diaz, and W. T. McAllister. 1993. T7 RNA polymerase mutants with altered promoter specificities. Proc. Natl. Acad. Sci. USA 90:3147–3151.

220. Robertson, E. S., L. A. Aggison, and A. W. Nicholson. 1994. Phosphorylation of elongation factor G and ribosomal protein S6 in bacteriophage T7-infected *Escherichia coli*. Mol. Microbiol. 11:1045–1057.

221. Robertson, E. S., and A. W. Nicholson. 1992. Phosphorylation of *Escherichia coli* translation initiation factors by the bacteriophage T7 protein kinase. Biochemistry 31:4822–4827.

222. Robins, W. P., and I. J. Molineux. Unpublished observations.

223. Roeder, G. S., and P. D. Sadowski. 1977. Bacteriophage T7 morphogenesis: phage-related particles in cells infected with wild-type and mutant T7 phage. Virology 76:263–285.

224. Rohwer, F., A. Segall, G. Steward, V. Seguritan, M. Breitbart, F. Wolven, and F. Azam. 2000. The complete genomic sequence of the marine phage Roseophage SIO1 shares homology with nonmarine phages. Limnol. Oceanogr. 45:408–418.

225. Rokyta, D., M. R. Badgett, I. J. Molineux, and J. J. Bull. 2002. Experimental genomic evolution: extensive compensation for loss of DNA ligase activity in a virus. J. Mol. Evol. 19:230–238.

226. Rong, M., B. He, W. T. McAllister, and R. K. Durbin. 1998. Promoter specificity determinants of T7 RNA polymerase. Proc. Natl. Acad. Sci. USA 95:515–519.

227. Rontó, G., M. M. Agamalyan, G. M. Drabkin, L. A. Feigin, and Y. M. Lvov. 1983. Structure of bacteriophage T7. Biophys. J. 43:309–314.

228. Rosa, M. D., and N. C. Andrews. 1981. Phage T3 contains an exact copy of the 28 base-pair phage T7 RNA polymerase promoter sequence. J. Mol. Biol. 147:41–53.

229. Rosenberg, A. H., K. Griffin, F. W. Studier, M. McCormick, J. Berg, R. Novy, and R. Mierendorf. 1996. T7Select phage display system: a powerful new protein display system based on bacteriophage T7, pp. 1–6. *In* Innovations—Newsletter of Novagen, Inc.

230. Rosenberg, A. H., S. S. Patel, K. A. Johnson, and F. W. Studier. 1992. Cloning and expression of gene *4* of bacteriophage T7 and creation and analysis of T7 mutants lacking the *4A* primase/helicase or the *4B* helicase. J. Biol. Chem. 267:15005–15012.

231. Rudolph, C., E. Freund-Mölbert, and S. Stirm, 1975. Fragments of *Klebsiella* bacteriophage No. 11. Virology 64:236–246.

232. Saito, H., and C. C. Richardson. 1981. Processing of mRNA by ribonuclease III regulates expression of gene 1.2 of bacteriophage T7. Cell 27:533–542.

233. Saito, H., and C. C. Richardson. 1981. Genetic analysis of gene *1.2* of bacteriophage T7: isolation of a mutant of *Escherichia coli* unable to support the growth of T7 gene *1.2* mutants. J. Virol. 37:343–351.

234. Schmitt, C. K. 1989. Incompatibility of T7 gene *1.2* with the F plasmid is the basis for F-mediated restriction of T7. PhD thesis, University of Texas at Austin.

235. Schmitt, C. K., P. Kemp, and I. J. Molineux. 1991. Genes *1.2* and *10* of bacteriophages T3 and T7 determine the permeability lesions observed in infected cells of *Escherichia coli* expressing the F plasmid gene *pifA*. J. Bacteriol. 173:6507–6514.

236. Schmitt, C. K., P. Kemp, and I. J. Molineux. 1995. Streptomycin and rifampicin-resistant mutants of *Escherichia coli* perturb exclusion of bacteriophage T7 by affecting synthesis of the F plasmid protein PifA. J. Bacteriol. 177:1589–1594.

237. Schmitt, C. K., and I. J. Molineux. 1991. Expression of gene *1.2* and gene *10* of bacteriophage T7 is lethal to F plasmid-containing *Escherichia coli*. J. Bacteriol. 173:1536–1543.

238. Schmitt, M. P., P. J. Beck, C. A. Kearney, J. L. Spence, D. DiGiovanni, J. P. Condreay, and I. J. Molineux. 1987. Sequence of a conditionally essential region of bacteriophage T3, including the primary origin of DNA replication. J. Mol. Biol. 193:479–495.

239. Scholl, D., S. Rogers, S. Adhya, and C. Merril. 2001. Bacteriophage K1-5 encodes two different tail fiber proteins, allowing it to infect and replicate on both K1 and K5 strains of *Escherichia coli*. J. Virol. 75:2509–2515.

240. Scholl, D., J. Kieleczawa, P. Kemp, J. Rush, C. C. Richardson, C. Merril, S. Adhya, and I. J. Molineux. 2003.

Genomic analysis of bacteriophages SP6 and K1-5, an estranged subgroup of the T7 supergroup. J. Mol. Biol. 335:1151–1171.

241. Serwer, P. 1976. Internal proteins of bacteriophage T7. J. Mol. Biol. 107:271–291.

242. Serwer, P. 1979. Fibrous projections from the core of a bacteriophage T7 capsid. J. Supramol. Struct. 58:235–243.

243. Serwer, P., S. A. Khan, S. J. Hayes, R. H. Watson, and G. A. Griess. 1997. The conformation of packaged bacteriophage T7 DNA: informative images of negatively stained T7. J. Struct. Biol. 120:32–43.

244. Serwer, P., R. H. Watson, and S. J. Hayes. 1982. Detection and characterization of agarose-binding, capsid-like particles produced during assembly of a bacteriophage T7 procapsid. J. Virol. 42:583–594.

245. Serwer, P., R. H. Watson, and S. J. Hayes. 1987. Multidimensional analysis of intracellular bacteriophage DNA: effects of amber mutations in genes *3* and *19*. J. Virol. 61:3499–3509.

246. Serwer, P., R. H. Watson, S. J. Hayes, and J. L. Allen. 1983. Comparison of the physical properties and assembly pathways of the related bacteriophages T7, T3, and ϕII. J. Mol. Biol. 170:447–469.

247. Shibata, H., H. Fujisawa, and T. Minagawa. 1987. Early events in DNA packaging in defined in vitro system of bacteriophage T3. Virology 159:250–258.

248. Shibata, H., H. Fujisawa, and T. Minagawa. 1987. Characterization of the bacteriophage T3 DNA packaging reaction in vitro in a defined system. J. Mol. Biol. 196:845–851.

249. Shin, I., J. Kim, C. R. Cantor, and C. Kang. 2000. Effects of saturation mutagenesis of the phage SP6 promoter on transcription activity, presented by activity logos. Proc Natl. Acad. Sci. USA 97:3890–3895.

250. Shinozaki, K., and T. Okazaki. 1977. RNA-linked nascent DNA pieces in T7 phage-infected *Escherichia coli* cells. Mol. Gen. Genet. 154:263–267.

251. Shinozaki, K., and T. Okazaki. 1978. T7 gene 6 exonuclease has RNase H activity. Nucleic Acids Res. 5:4245–4261.

252. Silberstein, S., M. Inouye, and F. W. Studier. 1975. Studies on the role of bacteriophage T7 lysozyme during phage infection. J. Mol. Biol. 96:1–11.

253. Simon, M. N., and F. W. Studier. 1973. Physical mapping of the early region of bacteriophage T7 DNA. J. Mol. Biol. 79:249–265.

254. Singleton, M. R., M. R. Sawaya, T. Ellenberger, and D. B. Wigley. 2000. Crystal structure of T7 gene *4* ring helicase indicates a mechanism for sequential hydrolysis of nucleotides. Cell 101:589–600.

255. Singleton, M. R., and D. B. Wigley. 2002. Modularity and specialization in superfamily 1 and 2 helicases. J. Bacteriol. 184:1819–1826.

256. Son, M., R. H. Watson, and P. Serwer. 1993. The direction and rate of bacteriophage T7 DNA packaging in vitro. Virology 196:282–289.

257. Sousa, R., Y. J. Chung, J. P. Rose, and B. C. Wang. 1993. Crystal structure of bacteriophage T7 RNA polymerase at 3.3Å resolution. Nature 364:593–599.

258. Sousa, R., and R. Padilla. 1995. A mutant T7 RNA polymerase as a DNA polymerase. EMBO J. 14:4609–4621.

259. Sousa, R., D. Patra, and E. M. Lafer. 1992. Model for the mechanism of bacteriophage T7 RNAP transcription and termination. J. Mol. Biol. 224:319–334.

260. Stano, N. M., and S. S. Patel. 2002. The intercalating β-hairpin of T7 RNA polymerase plays an active role in promoter DNA melting and in stabilizing the melted DNA for efficient RNA synthesis. J. Mol. Biol. 315:1009–1025.

261. Steven, A. C., P. Serwer, M. E. Bisher, and B. L. Trus. 1983. Molecular architecture of bacteriophage T7 capsid. Virology 124:109–120.

262. Steven, A. C., and B. L. Trus. 1986. The structure of bacteriophage T7, pp. 1–35. *In* J. R. Harris and R. W. Horne (eds.) Electron Microscopy of Proteins, vol. 5: Viral Structure. Academic Press, New york.

263. Steven A. C, B. L. Trus, J. V. Maizel, M. Unser, D. A. Parry, J. S. Wall, J. F. Hainfeld, and F. W. Studier. 1988. Molecular substructure of a viral receptor-recognition protein. The gp17 tail-fiber of bacteriophage T7. J. Mol. Biol. 200:351–365.

264. Strome, S., and E. T. Young. 1978. Translational control of the expression of bacteriophage T7 gene *0.3*. J. Mol. Biol. 125:75–93.

265. Strome, S., and E. T. Young. 1980. Chemical and functional quantification of gene *0.3* messenger RNA during T7 infection. J. Mol. Biol. 136:417–432.

266. Strome, S., and E. T. Young. 1980. Translational discrimination against bacteriophage T7 gene *0.3* messenger RNA. J. Mol. Biol. 136:433–450.

267. Stroud, R. M., P. Serwer, and M. J. Ross. 1981. Assembly of bacteriophage T7. Biophys. J. 36:743–757.

268. Struthers-Schlinke, J. S., W. P. Robins, P. Kemp, and I. J. Molineux. 2000. The internal head protein gp16 of bacteriophage T7 controls DNA ejection from the virion. J. Mol. Biol. 301:35–45.

269. Studier, F. W. 1969. The genetics and physiology of bacteriophage T7. Virology 39:562–574.

270. Studier, F. W. 1972. Bacteriophage T7. Science 176:367–376.

271. Studier, F. W. 1973. Genetic analysis of non-essential bacteriophage T7 genes. J. Mol. Biol. 79:227–236.

272. Studier, F. W. 1973. Analysis of bacteriophage T7 early RNAs and proteins on slab gels. J. Mol. Biol. 79:237–248.

273. Studier, F. W. 1975. Gene *0.3* of bacteriophage T7 acts to overcome the DNA restriction system of the host. J. Mol. Biol. 94:283–295.

274. Studier, F. W. 1979. Relationships among different strains of T7 and among T7-related bacteriophages. Virology 95:70–84.

275. Studier, F. W. 1981. Identification and mapping of five new genes in bacteriophage T7. J. Mol. Biol. 153:493–502.

276. Studier, F. W., and R. Hausmann. 1969. Integration of two sets of T7 mutants. Virology 39:587–588.

277. Studier, F. W., and J. V. Maizel. 1969. T7-directed protein synthesis. Virology 39:575–586.

278. Studier, F. W., and N. R. Movva. 1976. SAMase gene of bacteriophage T3 is responsible for overcoming host restriction. J. Virol. 19:136–145.

279. Studier, F. W., A. H. Rosenberg, J. J. Dunn, and J. W. Dubendorff. 1990. Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol. 185:60–89.
280. Subramanya, H. S., A. J. Doherty, S. R. Ashford, and D. B. Wigley. 1996. Crystal structure of an ATP-dependent DNA ligase from bacteriophage T7. Cell 85:607–615.
281. Sugimoto, K., Y. Kohara, and T. Okazaki. 1987. Relative roles of T7 RNA polymerase and gene *4* primase for the initiation of T7 phage DNA replication in vivo. Proc. Natl. Acad. Sci. USA 84:3977–3981.
282. Summers, W. C. 1970. The process of infection with coliphage T7. IV. Stability of RNA in bacteriophage-infected cells. J. Mol. Biol. 51:671–678.
283. Sun, M., D. Louie, and P. Serwer. 1997. Formation and cleavage of a DNA network during in vitro bacteriophage T7 DNA packaging: light microscopy of DNA metabolism. Biochemistry 36:13018–13026.
284. Sun, M., D. Louie, and P. Serwer. 1999. Single-event analysis of the packaging of bacteriophage T7 DNA concatemers in vitro. Biophys. J. 77:1627–1637.
285. Tabor, S., H. E. Huber, and C. C. Richardson. 1987. *Escherichia coli* thioredoxin confers processivity on the DNA polymerase activity of the gene *5* protein of bacteriophage T7. J. Biol. Chem. 262:16212–16223.
286. Tabor, S., and C. C. Richardson. 1981. Template requirements for RNA primer synthesis by gene *4* protein of bacteriophage T7. Proc. Natl. Acad. Sci. USA 78:205–209.
287. Tahirov, T. H., D. Temiakov, M. Anikin, V. Patlan, W. T. McAllister, D. G. Vassylyev, and S. Yokoyama. 2002. Structure of a T7 RNA polymerase elongation complex at 2.9 Å resolution. Nature 420:43–50.
288. Tamanoi, F., H. Saito, and C. C. Richardson. 1980. Physical mapping of primary and secondary origins of T7 DNA replication. Proc. Natl. Acad. Sci. USA 77:2656–2660.
289. Temiakov, D., P. E. Mentesana, K. Ma, A. Mustaev, S. Borukhov, and W. T. McAllister. 2000. The specificity loop of T7 RNA polymerase interacts first with the promoter and then with an elongating transcript, suggesting a mechanism for promoter clearance. Proc. Natl. Acad. Sci. USA 97:14109–14114.
290. Temiakov, D., V. Patlan, M. Anikin, W. T. McAllister, S. Yokoyama, and D. G. Vassylyev. 2004. Structural basis for substrate selection by T7 RNA polymerase. Cell 116: 381–391.
291. Toth, E. A., Y. Li, M. R. Sawaya, Y. Cheng, and T. Ellenberger. 2003. The crystal structure of the bifunctional primase-helicase of bacteriophage T7. Mol. Cell 12:1113–1123.
292. Tseng, T. Y., D. N. Frick, and C. C. Richardson. 2000. Characterization of a novel DNA primase from the *Salmonella typhimurium* bacteriophage SP6. Biochemistry 39:1643–1653.
293. Tsuchida, S., H. Kokubo, M. Tasaka, and H. Fujisawa. 1996. DNA sequences responsible for specificity of DNA packaging and phage growth interference of bacteriophages T3 and T7. Virology 217:332–337.
294. Újvári, A., and C. T. Martin. 1996. Thermodynamic and kinetic measurements of promoter binding by T7 RNA polymerase. Biochemistry 35:14574–14582.
295. Újvári, A., and C. T. Martin. 1997. Identification of a minimal binding element within the T7 RNA polymerase promoter. J. Mol. Biol. 273:775–781.
296. Újvári, A. and C. T. Martin. 2000. Evidence for DNA bending at the T7 RNA polymerase promoter. J. Mol. Biol. 295:1173–1184.
297. Valpuesta, J. M., H. Fujisawa, S. Marco, J. M. Carazo, and J. L. Carrascosa. 1992. Three-dimensional structure of T3 connector purified from overexpressing bacteria. J. Mol. Biol. 224:103–112.
298. Valpuesta, J. M., N. Sousa, I. Barthelemy, J. J. Fernández, H. Fujisawa, B. Ibarra, and J. L. Carrascosa. 2000. Structural analysis of the bacteriophage T3 head-to-tail connector. J. Struct. Biol. 131:146–155.
299. VanLoock, M. S., Y.-J. Chen, X. Yu, and S. S. Patel. 2001. The primase active site is on the outside of the hexameric bacteriophage T7 gene *4* helicase-primase ring. J. Mol. Biol. 311:951–956.
300. Villemain, J., and R. Sousa. 1998. Specificity in transcriptional regulation in the absence of specific DNA binding sites: the case of T7 lysozyme. J. Mol. Biol. 281:793–802.
301. Walkinshaw, M. D., P. Taylor, S. S. Sturrock, C. Atanasiu, T. Berge, R. M. Henderson, J. M. Edwardson, and D. T. F. Dryden. 2002. Structure of Ocr from bacteriophage T7, a protein that mimics B-form DNA. Mol. Cell 9:187–194.
302. Wang, I., D. L. Smith, and R. Young. 200**0.** Holins: the protein clocks of bacteriophage infections. Annu. Rev. Microbiol. 54:799–825.
303. Wang, W.-F., X. Cheng, and I. J. Molineux. 1999. Isolation and identification of *fxsA*, a gene that can suppress F exclusion of bacteriophage T7. J. Mol. Biol. 293:485–499.
304. Wang, W.-F., W. Margolin, W., and I. J. Molineux. 1999. Increased synthesis of an *Escherichia coli* membrane protein suppresses F exclusion of bacteriophage T7. J. Mol. Biol. 293:501–512.
305. Washington, M. T., Rosenberg, A. H., K. Griffin, F. W. Studier, and S. S. Patel. 1996. Biochemical analysis of mutant primase/helicase proteins defective in DNA binding, nucleotide hydrolysis, and the coupling of hydrolysis with DNA unwinding. J. Biol. Chem. 271:26825–26834.
306. Watanabe, T., T. Takano, T. Arai, H. Nishida, and S. Sato. 1966. Episome-mediated transfer of drug resistance in *Enterobacteriaceae*. X. Restriction and modification of phages by *fi*$^-$ R factors. J. Bacteriol. 92:477–486.
307. Watson, J. D. 1972. Origin of concatemeric T7 DNA. Nature New Biol. 239:197–201.
308. White, J. H., and C. C. Richardson. 1987. Processing of concatemers of bacteriophage T7 DNA in vitro. J. Biol. Chem. 262:8851–8860.
309. White, J. H., and C. C. Richardson. 1987. Gene *18* protein of bacteriophage T7. J. Biol. Chem. 262:8845–8850.
310. White, J. H., and C. C. Richardson. 1988. Gene *19* of bacteriophage T7. Overexpression, purification and characterization of its product. J. Biol. Chem. 263:2469–2476.
311. Wurgler, S. M., and C. C. Richardson. 1990. Structure and regulation of the dGTP triphosphohydrolase from *Escherichia coli*. Proc. Natl. Acad. Sci. USA 87:2740–2744.

312. Yamada, Y., and D. Nakada. 1976. Early to late switch in bacteriophage development: no translational discrimination between T7 early and late mRNA. J. Mol. Biol. 100:35–46.

313. Yamada, Y., J. Silnutzer, and D. Nakada. 1978. Mutant of *Escherichia coli* which blocks T7 bacteriophage assembly: accumulation of short T7 DNA. J. Mol. Biol. 121:95–121.

314. Yamada, Y., J. Silnutzer, and D. Nakada. 1979. Accumulation of bacteriophage T7 head-related particles in an *Escherichia coli* mutant. J. Virol. 31:209–219.

315. Yamada, Y., P. A. Whitaker, and D. Nakada. 1974. Functional instability of T7 early mRNA. Nature 248:335–338.

316. Yamada, Y., P. A. Whitaker, and D. Nakada. 1974. Early to late switch in bacteriophage T7 development; functional decay of T7 early messenger RNA. J. Mol. Biol. 89:293–303.

317. Yamada, Y., P. A. Whitaker, and D. Nakada. 1975. Chemical stability of bacteriophage T7 early RNA. J. Virol. 16:1683–1687.

318. Yamagishi, M., H. Fujisawa, and T. Minagawa. 1985. Isolation and characterization of bacteriophage T3/T7 hybrids and their use in studies on molecular basis of DNA-packaging specificity. Virology 144:502–515.

319. Yin, J. 1993. Evolution of bacteriophage T7 in a growing plaque. J. Bacteriol. 175:1272–1277.

320. Yin, Y. W. and T. A. Steitz. 2002. Structural basis for the transition from initiation to elongation transcription in T7 RNA polymerase. Science 298:1387–1395.

321. Yin, Y. W., and T. A. Steitz. 2004. The structural mechanism of translocation and helicase activity in T7 RNA polymerase. Cell 116:393–404.

322. Yu, X., M. M. Hingorani, S. S. Patel, and E. H. Egelman. 1996. DNA is bound within the central hole to one or two of the six subunits of the T7 DNA helicase. Nature Struct. Biol. 3:740–743.

323. Yuzenkova, J., S. Nechaev, J. Berlin, D. Rogulja, K. Kuznedelov, R. Inman, A. Mushegian, and K. Severinov. 2003. Genome of *Xanthomonas oryzae* bacteriophage Xp10: an odd t-odd phage. J. Mol. Biol. 330:735–748.

324. Zavriev, S. K., and Z. M. Kochkina. 1986. Bacteriophages T3 and T7: transcription-dependent mechanism of the transport of phage DNA into the cell during infection. Molek. Biol. 20:328–334.

325. Zavriev, S. K., and M. F. Shemyakin. 1982. RNA polymerase-dependent mechanism for the stepwise T7 phage DNA transport from the virion into *E. coli.* Nucleic Acids Res. 10:1635–1652.

326. Zavriev, S.K., and S.M. Vorob'ev. 1983. Penetration of T7 DNA associated with its transcription during infection. Molek. Biol. 17:1048–1059.

327. Zhang, X. 1995. T7 RNA polymerase and T7 lysozyme: genetic, biochemical and structural analysis of their interaction and multiple roles in T7 infection. PhD thesis, SUNY, Stonybrook, New York.

328. Zhang X., and F. W. Studier. 1995. Isolation of transcriptionally active mutants of T7 RNA polymerase that do not support phage growth. J. Mol. Biol. 250:156–168.

329. Zhang X., and F. W. Studier. 1997. Mechanism of inhibition of T7 RNA polymerase by T7 lysozyme. J. Mol. Biol. 269:964–981.

330. Zillig, W., H. Fujiki, W. Blum, D. Janekovic, M. Schweiger, H.-J. Rahmsdorf, H. Ponta, and M. Kirsch-Kaufmann. 1975. In vivo and in vitro phosphorylation of DNA-dependent RNA polymerase of *Escherichia coli* by bacteriophage T7-induced protein kinase. Proc. Natl. Acad. Sci. USA 72:2506–2510.

331. Zinder, N. D. 1961. A bacteriophage specific for F^- *Salmonella* strains. Science 133:2069–2070.

21

Bacteriophage N4

KRYSTYNA M. KAZMIERCZAK
LUCIA B. ROTHMAN-DENES

Bacteriophage N4 is a lytic phage specific for *Escherichia coli* K-12 strains originally isolated from the sewers of Genoa (80). N4 is unique in: (i) the use of three different DNA-dependent RNA polymerases during its growth cycle, (ii) a virion-encapsidated RNA polymerase (N4 vRNAP) that is injected into the host cell upon infection, (iii) the use of single-stranded DNA binding proteins as transcriptional activators, (iv) the presence of 3′ extensions at each end of its linear genome, and (v) a lysis-inhibited infection cycle.

The Virion

The N4 virion particle, as visualized by electron microscopy, consists of an icosahedral head 70 nm in diameter connected by a base plate to a small noncontractile tail, and a number of short tail fibers originating from the junction between the head and the tail (figure 21-1A) (82). The virion particle is composed of a single linear double-stranded DNA molecule and, as determined by sodium dodecyl sulfate–polyacrylamide gel electrophoresis, 10 proteins (25). The major component of the virion is the coat protein, a 48 kDa polypeptide. Also present is a virion-encapsidated phage-coded, DNA-dependent RNA polymerase (vRNAP) that is responsible for transcription of phage early genes (24).

The Genome

The genome, which has been recently sequenced, is a linear, double-stranded DNA molecule 70.6 kbp in length with a G-C content of 41.3 moles % (R. Hendrix, personal communication). Direct repeats varying in length from 390 to 440 bp and 3′ extensions are present at the ends of the genome (58). The left end is unique with microheterogeneity at the 3′-terminus yielding primarily the 5- or 6-base 3′ protruding sequences 3′-CATAA or 3′-CATAAA (figure 21-2A). In contrast, one major and at least three minor discrete families of the right and exist, differing in length by 10 bp and giving rise to the variability in the length of the terminal repeats. Each of these families of ends has a microheterogeneity of length with 1- to 3-base 3′ extensions (figure 21-2B) (58). The N4 genome is resistant to cleavage by a wide range of restriction endonucieases (58, 90).

The genome encodes three tRNAs and 72 open reading frames (ORFs), 51 of which show no statistically significant similarity to protein sequences in the database. Five early genes are located in two clusters at the left end of the genome and encode, among others, a protein implicated in genome injection (ORF1; A. Demidenko, unpublished data), and proteins that comprise the N4 middle transcriptional machinery (ORF2, ORF15, and ORF16; see below). Genes transcribed by the middle transcriptional machinery make up the remainder of the left half of the genome. Eleven middle ORFs encoding small proteins (ORFs 4–14) are located between the early gene clusters. Surprisingly, a 32 kDa virion protein (ORF17) of unknown function is transcribed during the middle period. ORF18 is homologous, on both the sequence and functional level, to bacteriophage P22 gp17, which enables P22 to successfully infect *Salmonella* strains containing the Fels-2 prophage (65; N. Federova, unpublished data). Five middle ORFs (ORFs 19–21, 46, and 48) are similar to ORFs encoded by the Salmonella *typhi* HMC 2 plasmid (59). ORFs 24 and 25 are similar to adjacent ORFs found in the *Streptomyces coelicolor* A3, *Thermotoga maritima*, and *Deinococcus radiodurans* genomes. ORFs 33 and 34 encode proteins with strong sequence similarity to putative domains of T4 RIIA and RIIB. N4 is currently the only phage unrelated to the T4 phage family known to encode RII-like proteins. Middle gene products also include four proteins required for N4 DNA replication (that will be discussed later) and ORFs with sequence similarity to dCTP deaminase/dUTPase (ORF26) and Thy 1, thymidylate synthase complementing the protein (ORF 30). Two middle genes of unknown function are located at the right end of the genome (M. Hammer, unpublished data). Late ORFs are located in the right half of the genome. These include genes for the vRNAP and other virion proteins, as well as a putative "lysis cassette."

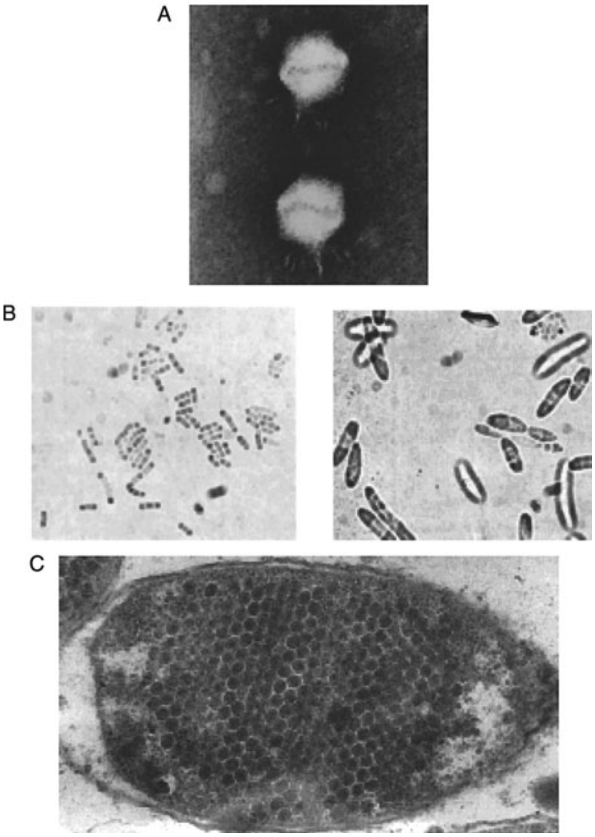


Figure 21-1 Phage N4 virions and infected bacteria. A: Electron micrograph of bacteriophage N4 virions stained with 2% phosphotungstic acid. Enlargement: ×175,000. Courtesy of M. Ohtsuki, University of Chicago. B: Phase-contrast microscopy of uninfected cells (left) and cells 2 hours after N4 infection (right). Courtesy of Dr. G. C. Schito, Institute of Microbiology, University of Genoa. C: Electron micrograph of an ultrathin section of *E. coli* K12S 4 hours after N4 infection. Enlargement: ×44,000. Courtesy of Dr. G. C. Schito, Institute of Microbiology, University of Genoa.

N4 Adsorption

N4 infection is initiated by phage attachment to at most five sites per bacterium. *E. coli* mutants to which N4 cannot absorb were isolated. The mutations map to four loci named *nfr*A, *nfr*B, *nfr*C, and *nfr*D (43). The *nfr*A and *nfr*B genes are located at 12.7 min on the *E. coli* linkage map. The *nfr*B gene lies upstream of the *nfr*A gene and the two overlap by 14 nucleotides. The *nfr*A gene encodes a 122 kDa outer membrane protein, while the *nfr*B gene encodes an 85 kDa inner membrane protein with three potential membrane-spanning regions (43, 44). Two ORFs with sequence similarity and similar organization to the *nfr*A and *nfr*B genes are also found in the *Ralstonia solanacearum* genome, though their functions are unknown (GenBank CAD 15982 and 17659, respectively). A plasmid expressing the NfrA and NfrB proteins confers sensitivity to N4 infection to *E. coli* B, which is normally resistant. The *nfr*C gene, located at 85.5 min within a cluster of genes involved in the synthesis of enterobacterial common antigen (ECA), is identical to *rff*E/*wec*B, and encodes a 42 kDa cytoplasmic protein, UDP-G1cNac 2-epimerase (42, 74). Mutations in *nfr*C affect neither the expression level or export of the NfrA protein, nor the expression or localization of NfrB in maxicells. The identity of *nfr*D, which maps to approximately 54.2 min, is unknown. It has been reported that *pel* mutants, which lack the II-PMan or II-Mman components of the mannose permease system required for λ phage DNA injection (20, 21), do not support N4 growth. N4 may interact with these proteins; however, N4 adsorption was not measured in these mutant strains (45). We have proposed that NfrA is the N4 receptor and that NrfB might provide a channel for injection of the genome and vRNAP (43). Since *nfr*C mutations do not affect the localization or amount of NfrA, we suspect that they might affect access to NfrA as a result of changes in ECA structure.

N4 Growth Cycle

Following phage adsorption, the genome and N4 vRNAP are injected into the host cell. Little is known about the mechanism of injection of vRNAP; the possible role of the N- or C-terminal domains of the vRNAP polypeptide in this process is under study (39). The process of genome injection is complex. The injection of the first 1000 bp of the genome, which occurs in the absence of post-infection protein synthesis, requires the N-terminal domain of the vRNAP polypeptide (A. Demidenko, unpublished data). Genome injection beyond the first 1000 bp requires the product of ORFI and vRNAP transcription from the three early promoters (see below). Host replication stops 3 minutes

A

```
          ▼
5'       TGGTGGGG-
3'  CATAAACCACCCC-
     ▲
```

B

```
                   ▼ ▼     ▼▼          ▼ ▼         ▼ ▼
5' -TCAGAAACAAATCTTAGACTTCTGCAAAGATGAACAAAACAAAGAATTGTTCTCT 3'
3' -AGTCTTTGTTTAGAATCTGAAGACGTTTCTACTTGTTTTGTTTCTTAACAAGAGA 5'
```

Figure 21-2 Structure of the ends of the N4 genome. A: Left end. B: Four families of sequences found at the right end; C predominates. Arrows mark the ends of the 3′ extensions.

after infection; an as yet unidentified early gene product must be responsible since host replication is not inhibited upon infection in the presence of chloramphenicol (72). The host genome is not degraded, and host messenger RNA and protein syntheses remain unaffected except for transcription of cAMP-dependent operons, which are shut off (72).

At least three products of early transcription are required for the synthesis of N4 middle transcripts, which begin to appear 3 minutes after infection. This transcriptional machinery is responsible for injection of the last 20 kbp of the N4 genome (A. Demidenko, unpublished data). Middle genes encode, in part, proteins required for N4 DNA replication. N4 DNA synthesis begins approximately 10–12 minutes after infection (73).

Late RNA synthesis, which begins 15 minutes after infection, is carried out by the *E. coli* σ^{70}-RNA polymerase (92) activated by the N4 single-stranded DNA binding protein (N4 SSB) (8). The first mature progeny are observed approximately 30 minutes post-infection and become localized to the cell poles (figure 21-1B) (79). Because N4 does not actively lyse the host cell, infected cells continue to grow, becoming enlarged and filled with a paracrystalline array of phage particles (figure 21-1C) (78). A yield of up to 3000 N4 particles per infected bacterium is obtained over a 3 hour period (78, 80).

N4 Transcription

The N4 genome is transcribed in three temporal stages—early, middle, and late— by three different DNA-dependent RNA polymerases (figure 21-3) (92). Both early and middle RNAs are transcribed with rightward polarity through the left half of the genome, while late transcription occurs through the right half of the genome with leftward polarity. Details of each of the three transcriptional stages follow.

Early Transcription

All known bacterial DNA viruses, except N4, use the host RNA polymerase to transcribe their early genes (67). N4 virions contain one or two copies of a phage-encoded RNA polymerase (vRNAP), which is injected with the phage genome upon infection and transcribes phage early genes during the first 5 minutes of infection (24). vRNAP activity in gently lysed cells is found to sediment with the DNA–membrane complex in sucrose gradients, making purification from infected cells difficult (22, 23). Therefore, vRNAP has been purified to homogeneity from virions and characterized (25).

Contrary to a previous report (61), the rifampicin- and streptolydigin-resistant vRNAP is a single polypeptide of estimated 320,000 molecular weight (25). Sequencing of the vRNAP gene revealed an ORF for a 3500 amino acid (aa) polypeptide that lacks extensive sequence similarity to either of the two known families of DNA-dependent RNA polymerases (39). However, vRNAP contains four short motifs (TxxGR, A, B, and C) characteristic of the family of T7-like single-subunit RNA polymerases. However, vRNAP contains four short motifs (TxxGR, A, B, and C) characteristic of the family of T7-like single-subunit RNA polymerases, which includes phge-encoded, mitochondrial and some chloroplast nuclear-encoded and linear plasmid enzymes (figure 21-4) (5, 39). To determine whether a smaller transcriptionally active domain exists within the polypeptide, vRNAP purified from virions was subjected to controlled trypsin proteolysis followed by a catalytic autolabeling assay (29, 33). Indeed, a stable and transcriptionally active 1106 aa long domain (mini-vRNAP), which possesses the same initiation, elongation, termination, and product displacement properties as full-length vRNAP, is located at the center of the vRNAP polypeptide (figure 21-4) (39). Mutational, biochemical, and phylogenetic analyses indicate that N4 mini-vRNAP is a highly evolutionarily divergent member of the single-subunit RNAP family (39).

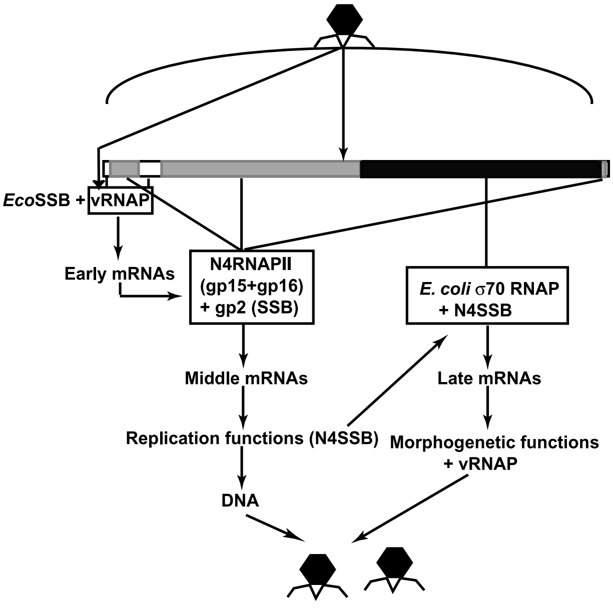

*Eco*SSB + vRNAP
Early mRNAs
N4RNAPII (gp15+gp16) + gp2 (SSB)
E. coli σ70 RNAP + N4SSB
Middle mRNAs
Late mRNAs
Replication functions (N4SSB)
Morphogenetic functions + vRNAP
DNA

Figure 21-3 Transcriptional program of the N4 genome. White blocks, early genes; gray blocks, middle genes; black block, late genes. The enzyme responsible for each transcriptional stage is indicated.

In vitro, vRNAP shows peculiar template specificity; it is inactive on linear double-stranded templates but transcribes denatured genomic N4 DNA or single-stranded promoter-containing DNA with in vivo specificity (26, 34). vRNAP recognizes three promoters, Pe1, Pe2, and Pe3, located at the left end of the genome (34). Promoter Pe1 is located within the terminal repeat and, therefore, a second copy exists at the right end of the genome (58). The three early promoters span positions −17 to +1 relative to the transcription start site and share blocks of conserved sequences and a set of inverted repeats centered at −11 (figure 21-5A) (34).

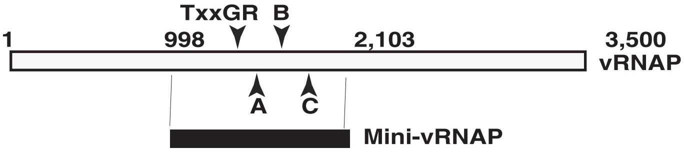


Figure 21-4 vRNAP and mini-vRNAP. Location of the transcriptionally active domain (mini-vRNAP) and the TxxGR, A, B, and C motifs in the vRNAP polypeptide. VRNAP stands for virion-encapsidated RNA polymerase.

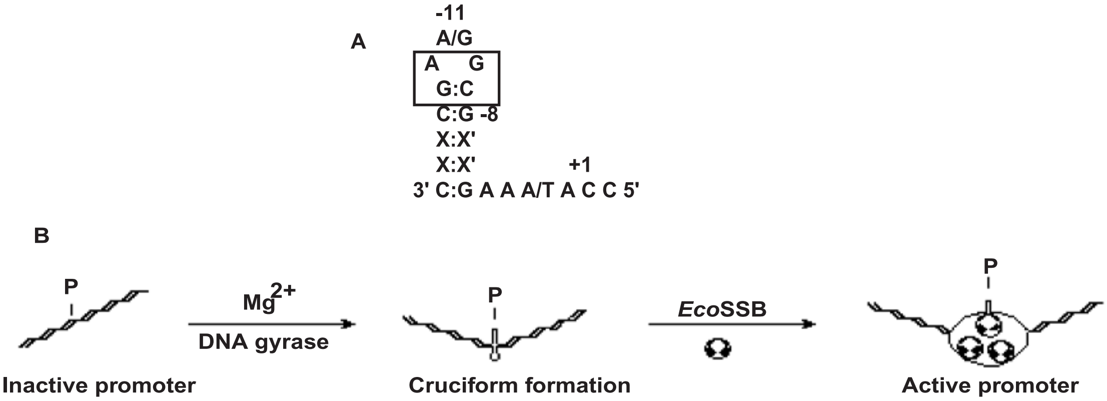


Figure 21-5 Phage N4 vRNAP promotion. A: Consensus sequence of the N4 virion-encapsidated RNA polymerase (vRNAP) promoters. Nonconserved positions in the inverted repeats are indicated by X:X′. The site of transcription initiation is indicated by +1. Boxed sequences are required for high hairpin stability and extrusion. B: Proposed model for the pathway of vRNAP polymerase–promoter utilization. The stoichiometry of *Eco* SSB in the activated complex is unknown.

All information necessary for promoter recognition and specific initiation is contained in the promoter template strand (27). Base changes in the nonconserved positions of the inverted repeats (X:X′) that disrupt potential base-pairing between the repeats leads to a severe decrease in promoter activity, suggesting that a hairpin structure is required for vRNAP-promoter recognition (27). Chemical and enzymatic footprinting confirmed that the inverted repeats, when present on single-stranded DNA templates, base-pair to form a hairpin structure with a 5 bp stem and 3 base loop (27). Run-off transcription experiments using templates containing mutant promoter sequences defined specific positions required for transcription (27). Determination of binding affinity (K_d) as well as UV crosslinking (312 nm) to deoxyoligonucleotide templates containing 5-iodo-deoxyuracil at positions within the promoter revealed that a purine at position −11 presented in the context of a hairpin loop is essential for promoter binding (E. Davydova, unpublished data). Regions of the transcriptionally active vRNAP domain that contact positions −11 and the hairpin stem were determined through crosslinking experiments using mini-vRNAP. Position −11 is contacted by amino acids present in the aa 83–200 interval (E. Davydova and K. M. Kazmierczak, unpublished data), while the hairpin stem is contacted by amino acids present in the aa 812–918 interval (figure 21-6) (E. Davydova, unpublished data). In contrast, promoter recognition in the distantly related T7 RNAP is conferred primarily by the specificity loop, located at aa 739–769 (figure 21-6) (6, 68, 71). In contrast to T7 RNAP, mini-vRNAP–promoter complexes are stable and resistant to dissociation by 2 M NaCl (17).

In vivo, early transcription is sensitive to *E. coli* DNA gyrase inhibitors, suggesting that the phage genome must undergo a structural change within the host cell in order to become a suitable template for vRNAP (26). However, no transcription was observed in vitro on promoter-containing plasmids of superhelical densities less than −0.071, twice the level of supercoiling found in the cell (16). A requirement for a second host factor, *Eco* SSB, was discovered upon N4 infection of a host strain carrying the conditional *ssb*-1 mutation; no N4 early transcription was observed in *ssb*-1 cells grown at the nonpermissive temperature (51). In vitro, *Eco* SSB stimulated specific transcription from promoters located on plasmids of physiological superhelical density (16, 51).

vRNAP transcripts generated from single-stranded templates are retained as RNA–DNA hybrids, indicating that vRNAP cannot displace the RNA product (26). Transcript release, as assayed by sensitivity to S1 or RNase H digestion, is observed in the presence of *Eco* SSB during

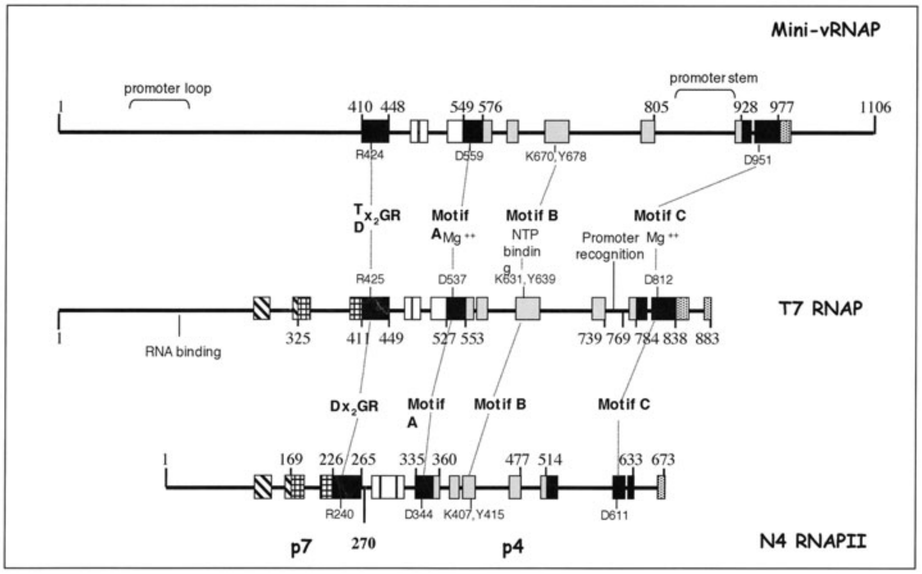


Figure 21-6 Sequence comparison of mini-vRNAP, T7 RNAP and N4 RNA polymerase II. Locations and sizes of conserved sequence blocks are schematized. Numbers indicate the known boundaries of structural domains in T7 RNA polymerase (36) and inferred boundaries in N4 mini-vRNA polymerase and N4 RNA polymerase II. Numbering in N4 RNA polymerase II reflects the positions of these blocks in the p7/p4 fusion polymerase sequence. Residue 270 corresponds to amino acid 1 of p4. Color coding is as follows: N-terminal domain, diagonally hatched; thumb, perpendicularly hatched; palm, black; palm insertion domain, white; fingers, gray; and foot module, pixels. The locations of polymerase motifs and catalytically important residues within them are indicated.

synthesis from a supercoiled double-stranded plasmid (51) or from promoter-containing oligonucleotide templates (18). Therefore, *Eco* SSB stimulates transcription through template recycling.

Eco SSB activation of vRNAP transcription is surprising because binding of SSBs to single-stranded DNA normally leads to melting of DNA secondary structures. However, enzymatic and chemical footprinting of *Eco* SSB on promoter-containing single-stranded DNA templates indicates that it is unable to melt the promoter hairpin upon binding, in contrast to other SSBs that do not activate transcription (fd gpV, T4 gp32, T7 gp2.5, N4 SSB) (28). The 177 aa *Eco* SSB polypeptide can be subdivided into a 120 aa N-terminal domain responsible for DNA binding (84), and a C-terminus consisting of a proline- and glycine-rich region followed by an acidic 10 aa tail highly conserved in eubacterial SSBs (49). Human mitochondrial SSB (*H*smt SSB) possesses sequence and structural similarity to the N-terminal domain of *Eco* SSB but lacks the conserved C-terminal domain (13, 83, 87). *H*smt SSB did not activate vRNAP transcription although it did not disrupt the promoter hairpin (18). Through the use of C-terminally truncated *Eco* SSB proteins and *H*smt SSB–*Eco* SSB C-terminal domain chimeras (12), we have shown that the last 10 aa of *Eco* SSB are essential for activation (18). We have been unable to detect vRNAP–*Eco* SSB interactions using purified proteins; however, it is possible that interactions only occur in the context of promoter–vRNAP or elongation complexes.

To explain the generation of a functional early promoter in vivo, we proposed that introduction of negative supercoiling by DNA gyrase induces the extrusion of a cruciform structure at the promoter that is invaded by *Eco* SSB to yield an "activated promoter" (figure 21-5B) (16). To test this model, plasmid minicircles comprised of N4 promoters Pe1 and Pe2 and their downstream ORFs were generated in vivo, isolated, and minicircle topoisomers of known superhelical densities prepared (54). The single-stranded DNA-specific probes, chloroacetaldehyde and mung bean nuclease, and T7 endonuclease, which recognizes DNA four-way junctions, were used to detect supercoiling-induced structural changes at the promoters. Supercoilng- and Mg^{2+} dependent cruciform extrusion was observed at physiological superhelical densities; surprisingly, only the nontemplate-strand hairpin loop is sensitive to single-stranded DNA-specific probes (15). A mutant promoter (P2flip), in which the bases within the loops of the template and nontemplate strand hairpins were switched, showed a corresponding switch in sensitivity to single-stranded DNA-specific probes. These results suggested that the loops of the two hairpins differ in conformation, a hypothesis supported by nuclear magnetic resonance data (10, 88; M. Kloster, unpublished data).

Mutational and extrusion analyses identified sequences essential for promoter cruciform extrusion. Extrusion requires a minimum stem length of 4 bp, provided that the stem is composed exclusively G:C base pairs, and that the loop and loop-closing base pair have the sequence 5′-C-GDA-G-3′ (where D = G, A, or T) (figure 21-5A) (15). This sequence has been shown to yield unusually stable hairpin (14, 36). These results, in addition to studies of promoter activity on single-stranded templates, reveal that the conserved sequences within the vRNAP promoters are important both for formation of the hairpin structure and for specific vRNAP–promoter contacts (15, 16, 27).

Analysis of the sequences downstream of ORF1 and ORF2, transcribed from Pe1 and Pe2 respectively, revealed eight or nine base palindromes followed by four or five thymidines. In vitro transcription of single-stranded or supercoiled, SSB-activated templates containing this region of the N4 genome led to the synthesis of defined transcripts initiating at Pe1 and Pe2 and terminating at t1 and t2, respectively (15, 27). Therefore, vRNAP recognizes eubacterial factor-independent-like termination signals.

Middle Transcription

Transcription of coliphage N4 middle messenger RNAs requires the activities of three early proteins: p17, p7, and p4 (73, 89, 92). Two of these proteins, p7 (30 kDa) and p4 (40 kDa), have been purified to homogeneity, and constitute a heterodimeric, rifampicin-resistant RNA polymerase, N4 RNA polymerase II (RNAPII) (90). However, this heterodimer does not bind to double-stranded, promoter-containing templates and transcribes promoter-containing single-stranded DNAs with low efficiency and no specificity (1, 90). RNAPII subunits p7 and p4 are tightly associated and copurify both from phage-infected cells (90) and, when overproduced, from plasmid-cloned genes (4). The p7 subunit was cloned with an N-terminal hexahistidine-tag; upon overproduction of the proteins, both hexahistidine-tagged p7 and native p4 were retained on a metal affinity column. Furthermore, the complex was resistant to dissociation with 1 M NaCl (4).

Sequencing of the genes encoding p7 and p4 identified ORFs 15 and 16, respectively (85). ORFs 15 and 16 display extensive sequence similarity to separate, nonoverlapping regions of T7 RNAP and other members of the single-subunit DNA-directed RNA polymerase family. RNAPII contains nearly perfect matches to four sequence motifs (DxxGR, A, B, and C) that are important for polymerase activity (figure 21-6) (85). The crystal structure of T7 RNAP resembles a cupped hand with thumb and finger subdomains that rise above the palm (38). The p7 subunit contains sequences corresponding to the N-terminal domain, the thumb, and a fragment of the palm in which the DxxGR motif is located, while p4 contains the remainder of the palm and the fingers subdomains, and all catalytically important residues within motifs A, B, and C (figure 21-6). This suggests that a gene fusion or gene splitting event occurred during the evolution of this class of polymerases. No sequence similarity to the N-terminal 264 aa of T7 RNAP was detected in RNAPII; in fact, the RNAPII N-terminus appears to be truncated by 156 aa relative to T7 RNAP (figure 21-6) (85). The T7 RNAP N-terminal 324 aa are involved in RNA binding, promoter recognition, unwinding and processive transcription (7, 30, 56, 75).

A consensus promoter sequence was derived by comparison of sequences upstream of six in vivo RNAPII transcription initiation sites (2). The recent availability of the complete N4 genome sequence led to the reexamination and identification of all middle transcription initiation sites (M. Hammer, unpublished). Through comparison of the activity of middle transcription initiation sites and genetic analysis of several middle promoters, a minimal N4 middle promoter consisting of an eleven nucleotide sequence directly upstream and inclusive of the initiating nucleotide (5′ tTttttTGA/GxG/T3′, +1 underlined) was identified. These data indicate a promoter comparable in size to that of yeast mtRNAP (3).

P17 is essential in vivo for N4 middle RNA synthesis (89). In wild-type phage-infected cells, RNAPII and p17 are tightly associated with an inner membrane/N4 DNA complex (23). Upon infection with phage containing an amber mutation in the gene encoding p17, RNAPII is found in the soluble fraction (89). P17 can be released from the DNA/membrane complex of N4-infected cells with 0.5 M NaCl. These results suggested that p17 might be a DNA binding protein that localizes RNAPII to DNA and confers promoter specificity. Yeast mitochondrial RNAP, which also belongs to the single-subunit family of DNA-dependent RNA polymerases, does not recognize its cognate promoters (40, 53); the specificity factor Mtf1p is required for promoter recognition (37, 76, 86). ORF2 encodes p17 (henceforth referred to as gp2), a 128 aa polypeptide that possesses no sequence similarity to proteins in the database (4). Gp2 was purified to homogeneity and assayed for binding to middle-promoter-containing DNA fragments. Gp2 was purified to homogeneity and assayed for binding to middle-promoter-containing DNA fragments. Gp2 does not form complexes with promoter-containing double-stranded DNA; however, it does bind to single-stranded DNA nonspecifically with high affinity ($K_d = 20$–60 nM) (4). Indeed, gp2 displays all properties characteristic of a single-stranded DNA binding protein (4).

In vitro, purified gp2 does not enable RNAPII to transcribe double-stranded, promoter-containing templates. However, it does activate nonspecific transcription by RNAPII on single-stranded templates (4). Gp2 activation is pronounced at low RNAPII concentrations and is absent at high RNAPII concentrations, suggesting that this protein might activate transcription through recruitment of polymerase to single-stranded templates (4). Gp2 and RNAPII bind to single-stranded DNA synergistically, suggesting the proteins do interact (4). Moreover, gp2 incubated with hexahistidine-tagged RNAPII before metal-affinity chromatography is retained with the polymerase on cobalt-agarose resin, further supporting an interaction between the proteins. After washing with 1 M NaCl, p4, p7 and gp2 are eluted in equimolar amounts (4). Other SSBs tested are not retained on the metal affinity column, indicating that gp2 and RNAPII interact specifically.

In vivo, expression of gp2 and N4 RNAPII are sufficient, for recognition when N4 middle promoters are present on plasmids (M. Hammer, unpublished data). The mechanism of promoter recognition and the role of gp2 in this process are under investigation. Moreover, whether gp2 simply recruits RNAPII to the promoter or plays an additional role in open complex formation remains to be determined.

Late Transcription

N4 late RNA synthesis, unlike early and middle transcription, is abolished by rifampicin (Rif) treatment in Rif-sensitive *E. coli* hosts but is unaffected in Rif-resistant hosts, indicating that *E. coli* RNA polymerase synthesizes late RNAs (92). No phage late messenger RNA synthesis is detected at the restrictive temperature in host strains carrying temperature-sensitive mutations in the α, β, β', and σ^{70} subunits of *E. coli* RNA polymerase, further confirming that the σ^{70} holoenzyme is responsible for phage late transcription (92).

All sites of late transcript initiation have been identified by S1 nuclease mapping (8) and/or primer extension (M. Hammer, unpublished data). Late promoter regions show weak homology to the *E. coli* σ^{70}-RNAP consensus -10 and -35 elements and possess no other blocks of conserved sequence. *E. coli* σ^{70} holoenzyme is inactive on linear DNA templates containing late promoters; however, it is active on supercoiled templates, raising the possibility that activation of late transcription might require a change in the topology of the DNA template (but see below) (8).

N4 late transcription begins shortly after the onset of phage DNA replication although concomitant or previous DNA replication is not required (8, 92). Mutant phage deficient in the N4-coded DNA polymerase or DNS, which are essential for N4 DNA replication, support late transcription albeit at a reduced rate (8). However, no late transcription occurs in cells infected with mutant phage deficient in N4 SSB, suggesting that this protein is required for both N4 replication and late transcription (8). Purified N4 SSB does not interact with double-stranded DNA, and binds to single-stranded DNA cooperatively (47). In vitro, N4 SSB activates *E. coli* σ^{70}-RNAP transcription at late promoters present on linear templates (8).

ORF45 encodes N4 SSB, a 265 aa protein with no sequence similarity to other single-stranded DNA binding proteins (9). A systematic mutational analysis of N4 SSB defined two functional domains separated by a linker region: a large N-terminal DNA binding domain and a short C-terminal domain implicated in protein–protein interactions (A. Miller, M. Choi, D. Wood, unpublished data). In vivo, mutations in N4SSB that affect single-stranded DNA binding (Y75A and Y128A) impair phage DNA replication and recombination but do not affect late transcription (55). Residues at the extreme C-terminus of N4 SSB are required for N4 DNA replication, recombination, and late transcription in vivo. Two mutant proteins Δ264–265:S260A and K264,265A, are proficient in both replication and recombination but deficient in transcriptional activation both in vivo, and in vitro, suggesting that residues S260, K264, and K265 play a specific role in this activation.

N4 SSB was retained on a metal affinity column containing either immobilized hexahistidine-tagged *E. coli* holoenzyme or core (55). Similar results were obtained with RNAP purified directly from cells or reconstituted from overexpressed subunits, indicating that no additional cellular proteins are required for this interaction (A. Miller, unpublished data). N4 SSB mutant proteins proficient in transcription activation but impaired in single-stranded DNA binding were retained on the RNAP column, while mutant proteins deficient in late transcriptional activation were not (55). Therefore, the observed interaction is specific and reflects contacts made between RNAP and N4 SSB that are necessary for transcriptional activation.

N4 phage cannot be propagated in an *E. coli rpoC*^{ts} strain, 397C, even at the permissive temperature (92). Specifically, *E. coli* 397C cannot support late transcription, as demonstrated by primer extension analysis (55). The rpoCts397C subunit has a 5 amino acid deletion at position 1354, followed by 23 out-of-frame amino acids; consequently, the 52 normally occurring C-terminal amino acids of β' are not present (11). Crosslinking experiments were performed to define the region of RNAP that interacts with N4 SSB. N4 SSB crosslinked to the 109 aa C-terminal peptide of the β' subunit (55). Therefore, N4 SSB activates *E. coli* σ^{70}-RNAP at N4 late promoters through a direct contact with the β' subunit (55).

Although activators can interact with RNAP in the absence of their cognate binding sites (52, 63), specific activation normally requires DNA binding specificity. N4 late promoters do not share any conserved sequences that might serve as an activator-binding site. How then does N4 SSB achieve specificity for activation at late promoters? Specificity may be the result of "indirect readout" of promoter sequences, in which N4 SSB preferentially recognizes a specific conformation that RNAP adopts only at N4 SSB dependent promoters. It could also be achieved kinetically, with N4 SSB affecting a step in transcription initiation that is limiting only at N4 SSB–RNAP complex may not exclusively transcribe from phage late promoters; such limited specificity might be acceptable for a viral activator acting late in a lytic cycle (55). Activators function by recruiting RNAP to the promoter (66) or by facilitating post-recruitment steps (e.g., open complex formation, promoter clearance) (3, 57, 64). Because N4 SSB activates transcription without binding to DNA, it most likely acts at post-recruitment steps.

N4 Replication

Host and Phage Functions Required for N4 DNA Replication

E. coli temperature-sensitive mutants defective in a number of replication functions were tested for their ability to support N4 growth and to permit incorporation of [^{3}H]thymidine into N4 DNA at the nonpermissive temperature (32, 69). Three host functions ribonucleotide reductase (*dnaF*), DNA ligase (*lig*), and DNA gyrase (*gyrB*)—were found to be required. *E. coli* cells defective in the polymerase activity of DNA polymerase I (*polA1*) were able to support phage growth, but cells defective in Pol I's 5′–3′ exonuclease activity (*polAex1*) could not, though [^{3}H]thymidine incorporation into phage DNA was observed in this strain. Analysis of replication products by alkaline sucrose gradient centrifugation suggests that the host exonuclease activity is required for processing of phage Okazaki fragments (32). The functions of the *E. coli dnaA*, *dnaB*, *dnaC*, *dnaE*, and *dnaG* genes are not required for N4 DNA synthesis, suggesting that the phage must encode the corresponding proteins required for replication initiation, priming, DNA unwinding, and polymerization. Five N4 gene products are required for in vivo N4 DNA replication: a DNA polymerase (N4 DNAP, ORF39), a single-stranded DNA binding protein (N4 SSB, ORF45), a 5′–3′ exonuclease (N4 5′–3′ EXO, ORF37), a protein of unknown function (DNS, ORF43) and vRNAP (ORF50) (32, 69).

An in vitro DNA replication system that utilizes a crude extract prepared fron N4-infected *E. coli polA1* cells was developed (70). This system requires exogenously added N4 DNA but shows little or no activity on other templates. The activities of the N4 DNAP, N4 SSB and N4 5′–3′ EXO are also required; therefore, these proteins were purified to apparent homogeneity by in vitro complementation (70).

N4 DNAP is monomeric in solution and has an apparent molecular weight of 87,000; this is lower than the molecular weight calculated based upon its sequence (888 aa, MW 101,262). N4 DNAP is homologous to members of the Pol I DNAP family. It absolutely requires a ribo- or deoxyriboprimer; small gapped DNAs serve as preferred templates. The N4 DNAP polypeptide has a strong Mg^{2+}-dependent 3′–5′ exonuclease activity that can be suppressed under polymerization conditions (46). The enzyme exhibits very high mismatch excision activity even at high nucleotide concentrations, a condition that normally favors polymerization over excision. Measurement of in vitro base substitution fidelity indicates that N4 DNAP is 16-fold more active in this assay than *E. coli* pol I Klenow fragment (46). As is true of other replicative DNA polymerases, N4 DNAP lacks 5′–3′ exonuclease and strand displacement activities and is nonprocessive, indicating the requirement for accessory factors. N4 SSB increases N4 DNAP's processivity 300-fold. It is unknown whether this stimulatory effect is the result of a specific DNA conformation elicited by N4 SSB binding or caused by SSB–DNAP interactions (47). Processivity factors commonly associated with replicative polymerases either form a structure that encircles the DNA or, in the case of T7 DNAP and its accessory factor thioredoxin, bind to DNAP and modify its structure so that it is able to encircle the DNA (19, 41). It is difficult to predict whether a clamp is encoded by the N4 genome; although clamps from different organisms display remarkable conservation of structure, they do not share sequence similarity (35).

As mentioned previously, N4 SSB is a 265 aa protein that binds to single-stranded DNA cooperatively and more tightly than to RNA. N4 SSB has a binding site size of 11 nucleotides, an intrinsic binding constant of $3.8 \times 10^4 M^{-1}$ and a cooperativity value (ω) of 300 when it binds to single-stranded DNA in 0.22 M NaCl at 37 °C (47). The in vivo concentration of N4 SSB is calculated to be 9–8 μM (47). In light of its binding site size and affinity, we predict that there is sufficient N4 SSB present in N4-infected cells to saturate all single-stranded DNA available during phage replication.

The N4 5′–3′ exonuclease has a denatured molecular weight of 45,000 and exists as a dimer in solution (31). Its preferred substrate is duplex DNA containing 3′ extensions (i.e., N4 DNA), which it degrades to 5′ mononucleotides by a distributive mechanism. It is inactive at nicks or gaps and on single-stranded DNA (31). These properties suggest a role for the exonuclease in recombination or replication of the ends of the N4 genome. The 439 aa exonuclease is encoded by ORF37, which displays striking sequence similarity to the T4-encoded DDA helicase. It is unknown whether the ORF37 product has intrinsic helicase activity. The phenotype of an exonuclease mutant suggests that the N4 enzyme serves a very different purpose in phage development than other phage-coded exonucleases. In vitro, no detectable exonuclease activity is found in extracts of nonsuppressor strains infected with N4 *exoam*D11 phage (31). In vivo, phage DNA synthesis proceeds normally in *exoam*D11 infected cells at 37 °C for 60 minutes, then stops. At 42°C, replication begins but ceases within a few minutes (D. Guinta, S. Spellman, unpublished data). The possible role of the exonuclease in N4 infection will be discussed below.

At least two additional N4-coded functions are required for N4 DNA replication in vivo: the ORF43 product (DNS) and vRNAP (32). ORF43 encodes a 715 aa protein with limited sequence similarity to an ORF located within the *S. aureus* spi pathogenicity island that has a putative role in replication. Analysis of phage encoding the vRNAP mutation *ts*150 revealed a role for vRNAP in N4 DNA replication. vRNAP purified from *ts*150 phage virions is sensitive to thermal inactivation, while enzyme purified from wild-type virions is temperature resistant (24). Shifting a culture of N4 *ts* 150-infected cells from the permissive to the restrictive temperature 25 minutes post-infection results in a rapid decrease in e rate of [^{3}H]thymidine incorporation into replicating phage DNA, thus suggesting that vRNAP plays a role in phage DNA replication. Interestingly, the *ts*150 mutation (H328Y) does not lie in the transcriptionally active central domain of the vRNAP polypeptide but instead is located in the amino-terminal domain (K.M.K., unpublished).

Mechanism of N4 DNA Replication

Replication in the in vitro system initiates at each end of the N4 genome and proceeds inward (70). Two-dimensional gel electrophoresis of restriction fragments from in vitro replicating N4 DNA molecules suggests that initiation occurs through hairpin priming of the genome's single-stranded ends (70). The in vitro replication system's dependence on the N4 5′–3′ exonuclease and the pattern of N4 DNA replication observed in N4*exoam*D11 infected cells at 42 °C might reflect a requirement for nucleotide removal from the genome's 5′ recessed ends so that hairpin formation and priming can occur. Electron microscopy of in vivo replicating N4 DNA revealed Y-shaped molecules and molecules with single-stranded tails. Strikingly, no replication bubbles were observed. These results strongly suggest that replication origins lie near or at the ends of the genome (48). However, the use of RNA primers for replication initiation has not been ruled out; indeed, N4 vRNAP's role in replication has not been elucidated. It is possible that vRNAP primes replication from promoter Pe1, located in the genome's terminal repeats (32). The accumulation of Okazaki-like fragments after infection of *polAex1* or *lig* mutants suggests discontinuous DNA synthesis (32), though an N4-coded primase activity has not been identified.

Longer than unit-length molecules were observed when replicating DNA was analyzed by electron microscopy (48). Restriction enzyme analysis identified a restriction fragment containing one copy of the terminal repeat flanked by the right and left ends of the genome (i.e., the joint fragment) (48). Furthermore, results of in vivo pulse-chase experiments show that label accumulates in the joint fragment and is chased into both ends of mature DNA molecules, strongly suggesting that concatemers of genomic DNA exist during N4 replication. It is likely that these originate through homologous recombination at the terminal repeats.

The 3′ extensions of the N4 genome are unusual; since all known DNA polymerases synthesize in a 5′ to 3′ direction, N4 must have a mechanism for generating these sequences. A model in which simple site-specific cleavage of concatemeric DNA during packaging yields the correct 3′ extensions at each end in one step is attractive, but by necessity implies that 50% of all synthesized DNA molecules would not be packaged into virions. Such a model, however, could account for the unit-length DNA lacking direct repeats observed in N4-infected ells (48). Moreover, it could explain the phenotype of *exoam*D11 infection at 37 °C. The N4 exonuclease may degrade alternate copies of genomic DNA in a concatemer in order to provide deoxynucleotides for continued DNA synthesis. Unlike other phage, N4 does not degrade the host chromosome during infection (32, 73) and consequently requires the activity of *E. coli* ribonucleotide reductase to provide a nucleotide pool (32). Perhaps a supplementary source of nucleotides is required late in phage infection, and in *exoam*D11 phage infection DNA synthesis stops as a result of nucleotide depletion.

N4 Genetics

Analysis of N4 development has been hindered by the lack of a genetic map; very high recombination levels have prevented the generation of a linkage map. *Hpa*I fragments of the N4 genome have been cloned into pBR322 (50). Two regions of the genome originating within *Hpa*I fragments M (ORFs 6, 7, and 8) and D (N4 SSB, ORFs 46 and 47) have been cloned only under tightly repressed conditions, indicating that they encode N4 functions lethal to the host. *Hpa*I M is transcribed early in infection and could encode the protein responsible for inhibiting host DNA replication. *Hpa*I fragment D encodes N4 SSB, which is lethal to the host even at low expression

Heat- and citrate-induced deletions defined nonessential regions of the N4 genome (50). Deletions were obtained with high frequency in *Hpa*I fragment A. The largest deletion spans a 6 kb region that includes ORFs 33 and 34, which encode proteins displaying sequence similarity to T4 RIIA and T4 RIIB (50; E. Stojkovic, unpublished data).

Morphogenesis

N4 does not encapsidate its DNA by full-head packaging but through the recognition of specific sequences at both ends of the genome, as indicated by our ability to isolate heat- and citrate-resistant N4 phage (60). As expected, phage genomes containing deletions have wild-type terminal sequences (50). The genes for known virion proteins have been identified, though the pathway of N4 virion assembly and DNA encapsidation is undefined. Recently, we addressed the possible role of the multifunctional vRNAP polypeptide in phage morphogenesis. The N- and C-terminal domains of vRNAP could potentially be important for encapsidation of vRNAP into virions, processing of DNA from concatemers and genome packaging into virions, or injection of vRNAP and the genome into the host cell.

Infection of nonsuppressing *E. coli* strains with phage containing amber mutations in vRNAP yields progeny phage apparently containing all virion proteins except vRNAP (I. Kaganman, unpublished data). These virions are morphologically indistinguishable from wild-type virions by electron microscopy, and contain one copy of N4 genomic DNA possessing wild-type ends, indicating that vRNAP is not involved in concatemer processing or DNA packaging.

N- and C-terminally truncated forms of the vRNAP polypeptide were expressed from plasmids during infection with vRNAP amber mutant phage; only polypeptides which contain the full vRNAP C-terminal domain were packaged into virions, indicating that this domain is necessary and sufficient for vRNAP encapsidation. Results of subsequent infection experiments indicate that the vRNAP's N-terminal domain is required for injection of the genome into the host cell (I. Kaganman, E. Davydova, A. Demidenko, unpublished data).

Lysis

N4 infection displays a lysis-inhibited phenotype (62, 80). However, treatment of N4-infected cells with chloroform induces lysis when added as early as 20 minutes after infection, suggesting that N4 encodes a lysin (E. Stojkovic, unpublished data). A late gene product must cause lysis because cells infected with N4 SSB mutant phage do not lyse upon chloroform addition.

Sequence analysis of N4 late genes identified ORF61 as a possible lysis gene due to its limited sequence similarity to a spore-cortex lytic enzyme from *Clostridium perfringens*. The 208 aa ORF61 protein does not display sequence similarity to other phage lysins but is homologous to hypothetical proteins encoded by the genomes of *Novosphingobium aromaticivorans*, *Salmonella typhimurium*, *Salmonella typhi*, *Neisseria meningitidis*, *Brucella melitensis*, *Mesorhizobium loti*, and *Rhodobacter capsulatus*.

ORF61 was cloned under the control of an inducible promoter. Cell lysis was detected 20 minutes after induction of gp61 synthesis, indicating that the ORF61 gene product does possess lysis activity and that this activity is able to access the cell peptidoglycan layer (E. Stojkovic, unpublished data). Gp61 is predicted to possess a positively charged N-terminus followed by a transmembrane domain. These features suggest that the protein is membrane bound with “N-in, C-out” topology. Proteins containing PhoA-fusions to the gp61 transmembrane domain or to its C-terminus displayed alkaline phosphatase activity, indicating that the PhoA domain was localized to the periplasm (E. Stojkovic, unpublished data). C-terminally hexahistidine-tagged recombinant gp61 purified to homogeneity degrades purified peptidoglycan, confirming that gp61 is a lysin (E. Stojkovic, unpublished data).

Analysis of muropeptides generated from gp61 digestion of *E. coli* peptidoglycan indicates that gp 61 is an N-acetyl muramidase; its homology to a number of hypothetical proteins defines a new family of lysins. The existence of an N4-encoded lysin raises the question of why N4 infection is lysis–inhibited (80). Inspection of the surrounding ORFs revealed that ORF61 is preceded by a small ORF (ORF62) containing two potential transmembrane domains, and followed by two overlapping ORFs (ORF60 and ORF60′), one of which (ORF60′) is out of frame. The organization of ORFs 60 and 60′ is reminiscent of the λ genes encoding Rz and Rz1. ORFs 63–60 are transcribed from a promoter located immediately upstream of a transcription terminator (E. Stojkovic, unpublished data). The roles of these ORF products in regulation of the gp61 lysin activity during N4 infection remain to be determined.

Prospects

We have presented our current understanding of bacteriophage N4. Sequencing of the genome has provided more questions than answers. What little we have learned already shows that N4 is unique among lytic phages. Important questions to be answered in the near future are:

1. What are the mechanisms of vRNAP and genome injection into the host? What is the exact role of the N-terminal domain of the vRNAP polypeptide in the process? What are the roles of NfrA and NfrB? Are additional host functions, i.e., chaperones, required for injection?
2. What is the structure of the central active domain of vRNAP (mini-vRNAP)? How does it interact with the promoter hairpin, and what conformational change occurs upon promoter binding to yield a salt resistant complex? What is the structure of the N4 early region in vivo that allows extrusion of the promoter hairpins? What is the in vivo structure of the activated early promoter and where is *Eco* SSB bound?
3. How is N4 RNAPII endowed with promoter specificity? What is the role of gp2 in this process? Do RNAPII accessory factors play other roles in addition to promoter recognition?
4. What is the mechanism by which N4 SSB activates *E. coli* σ^{70}-RNA polymerase at late promoters?
5. How is N4 replication initiated in vivo? What proteins are required for origin activation? What additional N4 functions are required at the replication fork? What is the role of vRNAP in N4 DNA replication?
6. How are the 3′ single-stranded ends of mature N4 DNA generated?
7. Where is vRNAP localized in virions? With what protein(s) does vRNAP interact in virions?
8. How is cell lysis regulated in N4-infected cells?

Acknowledgments

Work in our laboratory was supported by NIH grant AI 12575 K. M. K. was partially supported by United States Public Health Service Grant T32 GM07197.

References

1. Abravaya, K., and L. B. Rothman-Denes. 1989. In vitro requirements for N4 RNA polymerase II-specific initiation. J. Biol. Chem. 264:12695–12699.
2. Abravaya, K., and L. B. Rothman-Denes. 1989. N4 RNA polymerase II sites of transcription initiation. J. Biol. 211:359–372.
3. Cannon, W. V., M. T. Gallegos, and M. Buck. 2000. Isomerization of a binary sigma-promoter DNA complex by transcription activators. Nature Strut. Biol. 7:594–601.
4. Carter, R. H., A. A. Demidenko, S. H. Willis, and L. B. Rothman-Denes. 2003. The activator of N4 RNAPII is a single-stranded DNA binding protein. Genes Dev. 17:2334–2345.
5. Cermakian, N., T. M. Ikeda, P. Miramontes, B. F. Lang, M. W. Gray, and R. Cedergren. 1997. On the evolution of single-subunit RNA polymerases. J. Mol. Evol. 45:671–681.
6. Cheetham, G. M. T., D. Jeruzalmi, and T. A. Steitz. 1999. Structural basis for initiation of transcription from an RNA polymerase–promoter complex. Nature 399:80–83.
7. Cheetham, G. M. T., and T. A. Steitz. 2000. Insights into transcription: structure and function of single-subunit DNA-dependent RNA polymerases. Curr. Opin. Struct. Biol. 10:117–123.
8. Cho, N.-Y., M. Choi, and L. B. Rothman-Denes. 1995. The bacteriophage N4 single-stranded DNA binding protein (N4SSB) is the transcriptional activator of *E. coli* RNA polymerase at N4 late promoters. J. Mol. Biol. 246:461–471.
9. Choi, M., A. Miller, N-Y. Cho, and L. B. Rothman-Denes. 1995. Identification, cloning and characterization of the bacteriophage N4 gene encoding the single-stranded DNA binding protein: a protein required for phage replication, recombination and the late transcription. J. Biol. Chem. 270: 22541–22547.
10. Chou, S.-H., Y.-Y. Tseng, and B.-Y. Chu. 1999. Stable formation of a pyrimidine-rich loop hairpin in a cruciform promoter. J. Mol. Biol. 292:309–320.
11. Christie, G. E., S. B. Cale, L. A. Isaksson, D. J. Jin, M. Xu, B. Sauer, and R. Calendar. 1996. *Escherichia coli rpoC*397: a temperature sensitive C-terminal frameshift in the β′ subunit of RNA polymerase that blocks growth of bacteriophage P2. J. Bacteriol. 178:6991–6993.
12. Curth, U., J. Genschel, C. Urbanke, and J. Greipel. 1996. In vitro and in vivo function of the C-terminus of *Escherichia coli* single-stranded DNA binding protein. Nucleic Acids Res. 24:2706–2711.
13. Curth, U., C. Urbanke, J. Greipel, H. Gerberding, V. Tiranti, and M. Zeviani. 1994. Single-stranded-DNA-binding proteins from human mitochondria and *Escherichia coli* have analogous physicochemical properties. Eur. J. Biochem. 221:435–443.
14. Dai, X., M. Greizerstein, K. Nadas-Chinni, and L. B. Rothman-Denes. 1997. Supercoil-induced extrusion of a regulatory DNA hairpin. Proc. Natl. Acad. Sci. USA 94:2174–2179.
15. Dai, X., M. Kloster, and L. B. Rothman-Denes. 1998. Sequence-dependent extrusion of a small DNA hairpin at the N4 vRNAP promoters. J. Mol. Biol. 283:43–58.
16. Dai, X., and L. B. Rothman-Denes L. B. 1998. Sequence and DNA structural determinants of N4 virion RNA polymerase-promoter recognition. Genes Dev. 12:2782–2790.
17. Davydova, E. K., K. M. Kazmierczak, and L. B. Rothman-Denes. 2003. Bacteriophage N4-coded, virion-encapsulated DNA-dependent RNA polymerase, pp. 83–93. *In* S. Adhya, and S. Garges (eds.) Methods in Enzymology, vol. 370: RNA Polymerase and Associated Factors. Elsevier Academic Press, London.
18. Davydova, E. K. and L. B. Rothman-Denes. *Escherichia coli* single-stranded DNA-binding protein mediates template recycling during transcription by bacteriophage N4 virion RNA polymerase. Proc. Natl. Acad. Sci. USA 100:9250–9255.
19. Doublie, S., S. Tabor, A. M. Long, C. C. Richardson, and T. Ellenberger. 1998. Crystal structure of a bacteriophage T7 DNA replication complexed to a primer-template, a nucleoside triphosphate, and its processivity factor thioredoxin. Nature 391:251–258.
20. Elliott, J., and W. Arber. 1978. *E. coli* K-12 pel mutants, which block phage lambda DNA injection, coincide with *pts*M, which determines a component of a sugar transport system. Mol. Gen. Genet. 161:1–8.
21. Erni, B., B. Zanolari, and H. P. Kocher. 1987. The mannose permease of *Escherichia coli* consists of three different proteins. J. Biol. Chem. 262:5238–5247.
22. Falco, S. C., and L. B. Rothman-Denes. 1979. Bacteriophage N4-induced transcribing activities in *E. coli*. I. Detection and characterization in cell extracts. Virology 95:454–465.
23. Falco, S. C., and L. B. Rothman-Denes. 1979. Bacteriophage N4-induced transcribing activities in *Escherichia coli*. II. Association of the N4 transcriptional apparatus with the cytoplasmic membrane. Virology 95:466–475.
24. Falco, S. C., K. VanderLaan, and L. B. Rothman-Denes. 1977. Virion-associated RNA polymerase required for bacteriophage N4 development. Proc. Natl. Acad. Sci. USA 74:520–523.
25. Falco, S. C., W. Zehring, and L. B. Rothman-Denes. 1980. DNA-dependent RNA polymerase from bacteriophage N4 virions. Purification and characterization. J. Biol. Chem. 255:4339–4347.
26. Falco, S. C., R. Zivin, and L. B. Rothman-Denes. 1978. Novel template requirements of N4 virion RNA polymerase. Proc. Natl. Acad. Sci. USA 75: 3220–3224.
27. Glucksmann, M. A., P. Markiewicz, C. Malone, and L. B. Rothman-Denes. 1992. Specific sequences and a hairpin structure in the template strand are required for N4 virion RNA polymerase promoter recognition. Cell 70:491–500.
28. Glucksmann-Kuis, M. A., X. Dai, P. Markiewicz, and L. B. Rothman-Denes. 1996. *E. coli* SSB activates N4 virion

RNA polymerase promoters by stabilizing a DNA hairpin required for promoter recognition. Cell 84:147–154.

29. Grachev, M., T. Kolocheva, E. Lukhtanov, and A. Mustaev. 1987. Studies on the functional topography of *E. coli* RNA polymerase. Highly selective affinity labelling by analogues of initiating nucleotides. Eur. J. Biochem. 163:113–121.
30. Gross, L., W. J. Chen, and W. T. McAllister. 1992. Characterization of bacteriophage T7 RNA polymerase by linker insertion mutagenesis. J. Mol. Biol. 228:488–505.
31. Guinta, D., G. Lindberg, and L. B. Rothman-Denes. 1986. Bacteriophage N4 coded 5′–3′ exonuclease: purification and characterization. J. Biol. Chem. 261:10736–10743.
32. Guinta, D., J. Stambouly, S. C. Falco, J. K. Rist, and L. B. Rothman-Denes. 1986. Host and phage-coded functions required for coliphage N4 DNA replication. Virology 150:33–44.
33. Hartmann, G. R., C. Biebricher, S. J. Glaser, F. Grosse, M. J. Katzameyer, A. J. Lindner, H. Mosig, H.-P. Nasheuer, L. B. Rothmans-Denes, A. R. Schaffner, G. J. Schneider, K.-O. Stetter, and M. Thomm. 1988. Initiation of transcription: a general tool for affinity labeling of RNA polymerases by autocatalysis. Biol. Chem. Hoppe-Seyler 369:775–788.
34. Haynes, L. L., and L. B. Rothman-Denes. 1985. N4 virion RNA polymerase sites of transcription initiation. Cell 41:597–605.
35. Hingorani, M. M., and M. O'Donnell. 2000. Sliding clamps: A (tail)ored fit. Curr. Boil. 10:R25–R29.
36. Hirao, L., G. Kawai, S. Yoshizawa, Y. Nishimura, Y. Ishido, K. Watanabe, and K. Miura. 1994. Most compact hairpin turn structure exerted by a short DNA fragment, d(GCGAAGC) in solution: an extraordinary stable structure resistant to nucleases and heat. Nucleic Acids Res. 22:576–582.
37. Jang, S. H., and J. A. Jaehning. 1991. The yeast mitochondrial RNA polymerase specificity factor, MTFI, is similar to bacterial σ factors. J. Biol. Chem. 266:22671–22677.
38. Jeruzalmi, D., and T. A. Steitz. 1998. Structure of T7 RNA polymerase complexed to the transcriptional inhibitor T7 lysozyme. EMBO J. 17:4101–4113.
39. Kazmierczak, K. M., E. K. Davydova, A. A. Mustaev, and L. B. Rothman-Denes. 2002. The phage N4 virion RNA polymerase catalytic domain is related to single-subunit RNA polymerases. EMBO J. 21:5815–5823.
40. Kelly, J. L., and I. R. Lehman. 1986. Yeast mitochondrial RNA polymerase. Purification and properties of the catalytic subunit. J. Biol. Chem. 261:10340–10347.
41. Kelman, Z., and M. O'Donnell. 1995. Structural and functional similarities of prokaryotic and eukaryotic sliding clamps. Nucleic Acids Res. 23:3613–3620.
42. Kiino, D. R., R. Licudine, K. Wilt, D. Yang, and L. B. Rothman-Denes. 1993. A cytoplasmic protein, NfrC, is required for bacteriophage N4 adsorption. J. Bacteriol. 175:7074–7080.
43. Kiino, D. R., and L. B. Rothman-Denes. 1989. Genetic analysis of bacteriophage N4 adsorption. J. Bacteriol. 171:4595–4602.
44. Kiino, D. R., M. S. Singer, and L. B. Rothman-Denes. 1993. Two overlapping genes encoding membrane proteins required for bacteriophage N4 adsorption. J. Bacteriol. 175:7081–7085.
45. Kim, K., and O. J. Yoo. 1989. Two subunits of mannose permease, II-PMan and II-MMan, of *Escherichia coli* mediate coliphage N4 infection. Biochem. Int. 18:544–549.
46. Lindberg, G., J. K. Rist, T. Kunkel, A. Sugino, and L. B. Rothman-Denes. 1988. Purification and characterization of bacteriophage N4-induced DNA polyrnerase. J. Biol. Chem. 263:11319–11326.
47. Lindberg, G. J., S. C. Kowalczykowski, J. K. Rist, A. Sugino, A., and L. B. Rothman-Denes. 1989. Purification and characterization of the coliphage N4-coded single-stranded DNA binding protein. J. Biol. Chem. 264:12700–12708.
48. Lindberg, G. K., M. S. Pearle, and L. B. Rothman-Denes. 1988. N4 DNA replication in vivo and in vitro, pp. 130–139. *In* R. E. Moses, and W. C. Summers (eds.) DNA Replication and Mutagenesis. American Society for Microbiology, Washington, D.C.
49. Lohman, T., and M. E. Ferrari. 1994. *E. coli* single-stranded DNA binding protein: multiple DNA binding modes and cooperativities. Annu. Rev. Biochem. 63:527–570.
50. Malone, C., S. Spellman, D. Hyman, and L. B. Rothman-Denes. 1988. Cloning and generation of a genetic map of bacteriophage N4 DNA. Virology. 162:328–336.
51. Markiewicz, P., C. Malone, J. W. Chase, and L. B. Rothman-Denes. 1992. *E. coli* single-stranded DNA binding protein is a supercoiled-template dependent transcriptional activator of N4 virion RNA polymerase. Genes Dev. 6:2010–2019.
52. Martin, R. G., W. K. Gillette, N. I. Martin, and J. L. Rosner. 2002. Complex formation between activator and RNA polymerase as the basis for transcriptional activation by MarA and SoxS in *Escherichia coli*. Mol. Microbiol. 43:355–370.
53. Masters, B. S., L. L. Stohl, and D. A. Clayton. 1987. Yeast mitochondrial RNA polymerase is homologous to those encoded by bacteriophages T3 and T7. Cell 51:89–99.
54. Miller, A. A., X. Dai, M. Y. Choi, A. Glucksmann-Kuis, and L. B. Rothman-Denes. 1996. Single-stranded DNA binding proteins as transcriptional activators, pp. 9–19. *In* S. Adhya (ed.) Methods in Enzymology, vol. 174, part B: RNA Polymerase and Associated Factors. Academic Press, San Diego.
55. Miller, A. A., D. Wood, R. E. Ebright, and L. B. Rothman-Denes. 1997. RNA polymerase β′ subunit: a target of DNA binding-independent activation. Science 275:1655–1657.
56. Muller, D. K., C. T. Martin, and J. E. Coleman. 1988. Processivity of proteolytically modified forms of T7 RNA polymerase. Biochemistry 27:5763–5771.
57. Niu, W., Y. Kim, G. Tau, T. Heyduk, and R. H. Ebright. 1996. Transcription activation at Class II CAP-dependent promoters: two interactions between CAP and RNA polymerase. Cell 87:1123–1134.
58. Ohmori, H., L. L. Haynes, and L. B. Rothman-Denes. 1988. Structure of the ends of the coliphage N4 genome. J. Mol. Biol. 202:1–10.
59. Parkhill, J., G. Dougan, K. D. James, N. R. Thomson, D. Pickard, J. Wain, C. Churcher, K. L. Mungall, S. D. Bentley, M. T. Holden, M. Sebaihia, S. Baker, D. Basham, K. Brooks, T. Chillingworth, P. Connerton, A. Cronin, P. Davis, R. M. Davies, L. Dowd, N. White, J. Farrar, T. Feltwell,

N. Hamlin, A. Haque, T. T. Hien, S. Holroyd, K. Jagels, A. Krogh, T. S. Larsen, S. Leather, S. Moule, P. O'Gaora, C. Parry, M. Quail, K. Rutherford, M. Simmonds, J. Skelton, K. Stevens, S. Whitehead, and B. G. Barrell. 2001. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar *Typhi* CT18. Nature 413:848–852.

60. Parkinson, J. S., and R. J. Huskey. 1971. Deletion mutants of bacteriophage Lambda. I. Isolation and initial characterization. J. Mol. Biol. 56:369.
61. Pesce, A., C. Casoli, and G. C. Schito. 1978. Selectivity of transcription and structure of coliphage N4 virion-associated RNA polymerase. Biochem. Biophys. Res. Commun. 82:1040–1048.
62. Pesce, A., C. Casoli, and G. C. Schito. 1969. Factors in lysis-inhibition by N4 coliphage. Giorn. Microbiol. 17:119–129.
63. Pinkney, M., and J. G. Hoggett. 1988. Binding of the cyclic AMP receptor protein of *Escherichia coli* to RNA polymerase. Biochem. J. 250:897–902.
64. Popham, D. L., D. Szeto, J. Keener, and S. Kustu. 1989. Function of a bacterial activator protein that binds to transcriptional enhancers. Science 243:629–635.
65. Poteete, A. R. 1988. Bacteriophage P22, pp. 647–682. *In* R. Calendar (ed.) The Bacteriophages. Plenum Press, New York.
66. Ptashne, M., and A. Gann. 1997. Transcriptional activation by recruitment. Nature 386:569–577.
67. Rabussay, D., and E. P. Geiduschek. 1977. Regulation of gene action in development of lytic bacteriophages, pp. 1–196. *In* H. Fraenkel-Conrat, and R. R. Wagner (eds.) Comprehensive Virology, vol. 8. Plenum Press, New York.
68. Raskin, C. A., G. A. Diaz, and W. T. McAllister. 1993. T7 RNA polymerase mutants with altered promoter specificities. Proc. Natl. Acad. Sci. USA 90:3147–3151.
69. Rist, J. K., D. R. Guinta, A. Sugino, J., Stambouly, S. C. Falco, and L. B. Rothman-Denes. 1983. Bacteriophage N4 DNA replication, pp. 245–254. *In* N. R. Cozzarelli (eds.) Mechanisms of DNA Replication. Alan R. Liss, New York.
70. Rist, J. K., M. Pearle, A. Sugino, and L. B. Rothman-Denes. 1986. Development of an in vitro bacteriophage N4 DNA replication system. J. Biol. Chem. 261:10506–10510.
71. Rong, M., B. He, W. T. McAllister, and R. D. Durbin. 1998. Promoter specificity determinants of T7 RNA polymerase. Proc. Natl. Acad. Sci. USA 95:515–519.
72. Rothman-Denes, L. B., R. Haselkorn, and G. C. Schito. 1972. Selective shutoff of catabolite-sensitive host syntheses by bacteriophage N4. Virology 50:95–102.
73. Rothman-Denes, L. B., and G. C. Schito. 1974. Novel transcribing activities in N4 infected *Escherichia coli* Virology 60:65–72.
74. Sala, R. F., P. M. Morgan, and M. E. Tanner. 1996. Enzymatic formation and release of a stable glycal intermediate: the mechanism of the reaction catalyzed by UDP-*N*-acetylglucosamine 2-epimerase. J. Am. Chem. Soc. 118:3033–3034.
75. Sastry, S., and B. M. Ross. 1998. RNA-binding site in T7 RNA polymerase. Proc. Natl. Aced. Sci. USA 95:9111–9116.
76. Schinkel, A. H., M. J. A. Groot Koerkamp, E. P. W. Touw, and H. F. Tabak. 1987. Specificity factor of yeast mitochondrial RNA polymerase. J. Biol. Chem. 262:12785–12791.
77. Sinkel, A. H., M. J. A. Groot Koerkamp, M. H. Stuiver, G. T. J. Van der Horst, and H. F. Tabak. 1987. Effect of point mutations on in vitro transcription from the promoter for the large ribosomal RNA gene of yeast mitochondria. Nuc. Acids. Res. 15:5597–5612.
78. Schito, G. C. 1967. Intracellular crystallization of bacteriophage N4. Virology 32:723–725.
79. Schito, G. C. 1974. Development of coliphage N4: ultrastructural studies. J. Virol. 13:186–196.
80. Schito, G. C., A. M. Molina, and A. Pesce. 1967. Lysis and lysis inhibition with N4 coliphage. Giorn. Microbiol. 15:229–244.
81. Schito, G. C., A. M. Molina, and A. Pesce. 1965. Un nuovo batteriofago attivo sul ceppo K12 di E. coli I. Caractteristiche biologiche. Boll. Inst. Sieroter. Milanese 44:329–332.
82. Schito, G. C., G. Rialdi, and A. Pesce. 1966. Biophysical properties of N4 coliphage. Biochim. Biophys. Acta 129:482–490.
83. Webster, G., J. Genschel, U. Curth, C. Urbanke, C. Kang, and R. Hilgenfeld. 1997. A common core for binding single-stranded DNA: structural comparison of the single-stranded-DNA-binding proteins (SSB) from *E. coli* and human mitochondria. FEBS Lett. 411:313–316.
84. Williams, K. R., E. K. Spicer, M. LoPresti, R. A. Guggenheimer, and J. W. Chase. 1983. Limited proteolysis studies on the *E. coli* single-stranded DNA binding protein: evidence for a functionally homologus domain in both the *E. coli* and T4 DNA binding proteins. J. Biol. Chem. 258:3346–3355.
85. Willis, S. H., K. M. Kazmierczak, R. H. Carter, and L. B. Rothman-Denes. 2002. N4 RNA polymerase II, a heterodimeric RNA polymerase with homology to the single-subunit family of RNA polymerases. J. Bacteriol. 184:4952–4961.
86. Winkley, C. S., M. J. Keller, and J. A. Jaehning. 1985. A multicomponent mitochondrial RNA polymerase from *Saccharomyces cerevisiae.* J. Biol. Chem. 260:14214–14223.
87. Yang, C., U. Curth, C. Urbanke, and C. Kang. 1997. Crystal structure of the human mitochondrial single-stranded DNA binding protein at 2.4 Å resolution. Nature Struct. Biol. 4:153–157.
88. Yoshizawa, S., G. Kawai, K. Watanabe, K. Miura, and I. Hirao. 1997. GNA trinucleotide loop sequences producing extraordinarily stable DNA hairpins. Biochemistry 36:4761–4767.
89. Zehring, W. A., S. C. Falco, C. Malone, and L. B. Rothman-Denes. 1983. Bacteriophage N4-induced transcribing activities in *E. coli.* III. A third cistron required for N4 RNA polymerase II activity. Virology 126:678–687.
90. Zehring, W. A., and L. B. Rothman-Denes. 1983. Purification and characterization of coliphage N4 RNA polymerase II activity from infected cell extracts. J. Biol. Chem. 258:8074–8080.
91. Zivin, R., C. Malone, and L. B. Rothman-Denes. 1980. Physical map of coliphage N4 DNA. Virology 104:205–218.
92. Zivin, R., W. A. Zehring, and L. B. Rothman-Denes. 1981. Transcriptional map of bacteriophage N4. Location and polarity of N4 RNAs. J. Mol. Biol. 152:335–356.

22

Phage ϕ29 and its Relatives

MARGARITA SALAS

Phage ϕ29 and the related phages PZA, PZE, ϕ15, BS32, B103, Nf, M2Y, and GA-1 are lytic phages that belong to the *Podoviridae* family. Most of them infect *Bacillus subtilis* and other related species such as *Bacillus amyloliquefaciens*, *Bacillus pumilus,* and *Bacillus licheniformis*. The host range of phage GA-1 differs from those of the other phages. Thus, GA-1 infects the *Bacillus* strain G1R. The ϕ29-related phages have been classified in three groups based on serological properties, peptide maps, DNA physical maps, and DNA sequences (11, 90, 153). The first group includes phages ϕ29, PZA, PZE, ϕ15, and BS32; the second includes phages B103, Nf, and M2Y; and the third group contains phage GA-1 as the only member.

The genome of the ϕ29-related phages, which are the smallest of the *Bacillus* phages isolated so far, consists in a linear double-stranded DNA of 19–21 kb with a phage-encoded protein, named terminal protein (TP), covalently linked to each 5′ end. The complete DNA sequence of phages ϕ29 (19,285 bp), PZA (19,366 bp), B103 (18,630 bp), and GA-1 (21,129 bp) are known (90, 109, 112, 149). Phage ϕ29 and its relatives have a short, inverted terminal repeat, six nucleotides long (5′-AAAGTA) for ϕ29, PZA, ϕ15, and B103 DNAs, eight nucleotides long (5′-AAAGTAAG) for Nf and M2Y DNAs, and seven nucleotides long (5′-AAATAGA) for GA-1 DNA. The remainder of the DNA sequence of phages ϕ29 and PZA is very similar, and different from that of phage B103. The sequence of GA-1 DNA is less related to that of the other phages.

The phage ϕ29 particle consists of a prolate head (54 nm long and 42 nm wide), and a neck/tail region (44 nm long) (141). The head, formed by protein p8, contains fibers formed by protein p8.5. The neck consists of an upper collar or connector (p10), required for head assembly, and a lower collar (p11) to which 12 spindle-shaped appendages are attached (p12*). These appendages are required for adsorption of the phage to the bacterial cell wall. The tail is formed by protein p9.

In this chapter I describe the molecular mechanisms involved in DNA replication, regulation of transcription, and phage morphogenesis, most of which have been elucidated by studying phage ϕ29.

Genetic and Transcriptional Organization of ϕ29 and its Relatives

As shown in figure 22-1, the genetic and transcriptional maps of phages ϕ29, B103, and GA-1, representative of groups I, II, and III, respectively, are similarly organized. Early genes are located at the two DNA ends and they are transcribed from right to left, whereas late genes, located at the center of the genome, are transcribed from left to right. In the three phages, early genes 2, 3, 5, 6, 16.7, and 17 encode similar proteins required for DNA replication, and gene 4 encodes a protein involved in transcription regulation. Protein p6 also has a role in transcription regulation, in addition to its requirement for DNA replication. Gene 1, also involved in DNA replication, is present only in phage ϕ29. There are other open reading frames (ORFs) in the three phages whose function is still unknown. It should be noticed that, in accordance with its larger size, GA-1 DNA has more ORFs in the region corresponding to early genes than ϕ29 and B103 DNAs. The three phage DNAs contain a region at their left end encoding a small RNA, named pRNA, required for packaging of the phage DNA. The late genes of the three phages are very similar, except that phage GA-1 lacks gene 8.5, coding for the head fibers, which in ϕ29 is a dispensable protein. The functions of the different genes of phage ϕ29, their size, and the percentage similarity with the corresponding genes of phages B103 and GA-1 are given in table 22-1. In all cases, late transcription occurs from a single promoter, named A3. The major early promoters, A2b, A2c, and C2, are also conserved in the three phage DNAs. The A1 promoter (A1b in phage GA-1), responsible for the synthesis of pRNA, is also conserved in the three phages. The remaining, minor promoters are specific to each phage (see figure 22-1).

Table 22-1 Genes of phage ϕ29 and their function

Gene	Function	No. of amino acids	Mol. mass (kDa)	Similarity (%)[a]		
				ϕ29/B103	ϕ29/GA-1	B103/GA-1
1	DNA replication	89	10.3	–	–	–
2	DNA polymerase	572	65.2	88.5	67.3	68.0
3	Terminal protein	266	31.1	74.1	51.7	53.6
4	Transcriptional regulator	125	15.1	83.2	57.6	58.4
5	ssDNA binding protein	124	13.4	73.8	37.9	45.1
6	dsDNA binding protein	104	12.0	65.0	52.7	54.8
7	Scaffolding protein	98	11.2	75.5	48.0	38.6
8	Major head protein	448	49.7	90.4	68.5	69.3
8.5	Head fiber protein	280	29.6	63.6	–	–
9	Tail protein	599	67.7	75.3	58.3	58.5
10	Connector (upper collar protein)	309	35.9	84.4	63.7	63.7
11	Lower collar protein	293	33.6	77.8	51.6	50.9
12	Preneck appendage protein	854	92.4	79.9	37.8	37.5
13	Morphogenesis (tail assembly)	365	41.0	82.7	67.0	67.3
14	Holin	131	14.9	87.1	51.9	52.3
15	Peptidoglycan hydrolase	258	26.9	74.8	37.2	40.6
16	ATPase, DNA encapsidation	332	38.9	86.3	67.5	68.8
16.7	DNA replication, membrane protein	130	15.2	68.5	48.5	47.7
17	DNA replication	167	19.4	48.1	44.8	37.9

Data taken from Salas (121), and Meijer et al. (90).

ds, double-stranded; ss, single stranded.

[a] For the calculation of the percent similarity, the following amino acids were considered to be conservative: L, I, V, A, and M; F, Y, and W; K and R; D and E; Q and N; S and T.

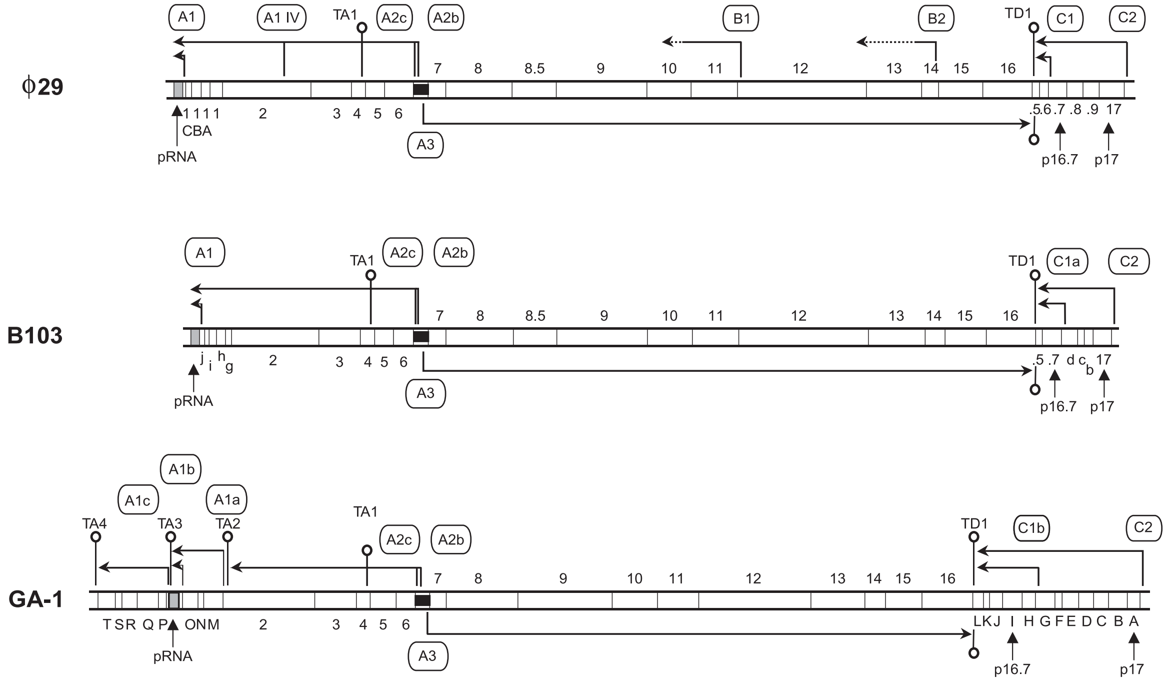

ϕ29
A1
A1 IV
TA1
A2c
A2b
B1
B2
TD1
C1
C2
7 8 8.5 9 10 11 12 13 14 15 16
1 1 1 1 2 3 4 5 6
CBA
pRNA
A3
.5 6 .7 .8 .9 17
p16.7 p17
B103
A1
TA1
A2c
A2b
TD1
C1a
C2
7 8 8.5 9 10 11 12 13 14 15 16
j i h g 2 3 4 5 6
pRNA
A3
.5 .7 d c b 17
p16.7 p17
GA-1
A1b
A1c
A1a
TA4
TA3
TA2
TA1
A2c
A2b
TD1
C1b
C2
7 8 9 10 11 12 13 14 15 16
T S R Q P O N M 2 3 4 5 6
pRNA
A3
L K J I H G F E D C B A
p16.7 p17

Figure 22-1 Genetic and trasncriptional maps of ϕ29, B103, and GA-1 DNA. The maps are aligned with respect to the A2c, A2b, and A3 promoters. The direction of transcription and length of the transcripts are indicated by arrows. The positions of genes 16.7 and 17 are indicated. ORFs 16.9, 16.8, 16.6, and 16.5 are indicated by the numbers .9, .8, .6, and .5, respectively. Transcriptional terminators are indicated by stem-loop structures. The grey box indicates the DNA region encoding the pRNA, and the black box indicates the region containing the early promoters A2c and A2b, and the late promoter A3. The figure is adapted from Meijer et al. (90).

All early promoters of the three phages have the consensus sequences at the −10 (TATAAT) and −35 (TTGACA) regions separated by a spacer of 17–18 nucleotides, except the ϕ29 minor promoter A1IV that lacks the −10 hexamer. In the three phages, the late A3 promoter has a −10 hexamer (TATAAT) but lacks the −35 one. Most of the promoters of the three phages have the so-called −10 extended region or TG motif (TGN-10). The involvement of the TG motif in promoter strength has been studied for the ϕ29 promoters A1, A2c, and A3. Mutation of the TG motif impaired the binding of σ^A RNA polymerase (RNAP) to each promoter (30), suggesting that the TG motif provides additional contact sites for the RNAP.

Transcriptional Regulation

Early Promoter C2: Regulation by Protein p6

The activity of the early promoter C2 of phage ϕ29 decreases very soon after infection (102). Protein p6, an early double-stranded DNA binding protein expressed in high amounts and required for the initiation of ϕ29 DNA replication, was shown to be responsible for in vivo and in vitro repression of promoter C2 (3, 152). Protein p6 binds to the ϕ29 DNA ends forming a nucleoprotein complex (see later). Nonetheless, the complex formed at the right DNA end does not occlude promoter C2 to the RNAP. Protein p6 binding alters the structure of the transcriptional complex and affects the stability of the closed complex (31).

The GA-1 C2 promoter is also expressed only early after infection. The GA-1 protein p6 inhibits in vitro transcription from this promoter (77).

Early Promoters A2b and A2c and Late Promoter A3: Regulation by Proteins p4 and p6

As already mentioned, the late A3 promoter lacks the −35 hexamer, so that the RNAP cannot bind to it. Instead, the sequence 5′-CTTTTT-15bp-AAAATG-3′ forms a recognition site for protein p4 at position −82 from the transcription start site. This region has an intrinsic curvature of 45° that increases up to 80–85° on p4 binding (117). A kinetic analysis of the activation process, as well as band-shift and DNase I footprinting assays, showed that the main role of protein p4 is to stabilize the binding of RNAP to the promoter as a closed complex (108) through a specific interaction between protein p4 and RNAP. Thus, protein p4 mutants affecting residue Arg120 bind to DNA efficiently but they have a reduced ability to interact with RNAP and to activate transcription (96). The protein p4 C-end is also known to participate in the maintenance of DNA bending. Reduction of bending was significant when two basic residues were simultaneously mutated in any combination. Therefore, the C-end of p4 participates both in DNA bending and in transcription activation. Considering that protein p4 is a dimer in solution and binds to DNA as a tetramer, a model was proposed in which two p4 subunits would be in close contact with the DNA, interacting with the DNA backbone and bending it. The other two subunits of p4 would be free to interact with RNAP through the surface centered around Arg120. On the other hand, protein p4 stabilizes the purified α-subunit of *B. subtilis* at the A3 promoter through an interaction that is dependent on p4 residue, Arg120. To localize the contact site of protein p4 at the RNAP, deletions at the C-end of the α-subunit, lacking the 15, 37 or 59 residues, were obtained. RNAPs reconstituted with any of the deletions could not recognize the late A3 promoter in the presence of p4. DNase I footprinting assays showed that protein p4 was unable to stabilize the mutant RNAPs at the A3 promoter (95). Therefore, the interaction between p4 and RNAP, which stabilizes the latter at promoter A3 to activate transcription, is maintained between the protein p4 region, containing residue Arg120 and the C-terminal domain (CTD) of the RNAP α-subunit.

The protein p4 binding site at the late A3 promoter overlaps with the −35 region of the early promoter A2b. In vitro studies have shown that p4 represses the A2b promoter by excluding the RNAP from the latter, directing it to the late promoter. In addition, the curvature induced by p4 binding impairs transcription from promoter A2b (116). Activation of the A3 promoter of phage Nf, which belongs to the same group as phage B103, responds to the ϕ29 protein p4 in a similar way to that described for the ϕ29 promoter A3 (107). In the case of phage GA-1, the corresponding p4 protein, named $p4_G$, binds to a site upstream of the late A3 promoter that overlaps with the early A2b promoter, preventing the binding of the RNAP to the latter and repressing transcription. However, binding of p4 to its site upstream the late A3 promoter did not activate transcription. Promoter A3 was expressed efficiently in vitro in both the absence or the presence of $p4_G$. Nonetheless, promoter A3 of GA-1 was not active in vivo when protein synthesis was inhibited. These results lead to the suggestion that the GA-1 A3 promoter may be repressed in vivo by a host protein and that protein $p4_G$ may act as an antirepressor, allowing expression of promoter A3 at late infection times (78).

Protein p4 also represses the early promoter A2c by binding to DNA immediately upstream from RNAP in a way that does not hinder polymerase binding. On the contrary, the two proteins bind cooperatively to DNA. Upstream from the A2c promoter there are two protein p4 binding sites, named site 1 and site 2. Protein p4 by itself binds to site 1 with low efficiency. In the presence of RNAP, protein p4 is displaced from site 1 to site 2, increasing the efficiency of binding (100). In the presence of p4, RNAP can form an initiated complex at the A2c promoter that generates short abortive transcripts, but cannot leave the promoter (103), leading to transcription repression by a mechanism different from the one that occurs at the adjacent A2b promoter. Thus, repression of the A2c promoter occurs at the step of promoter clearance, while that of promoter A2b takes place by exclusion of the RNAP from the promoter due to the binding of protein p4 to its recognition site upstream from the A3 promoter (site 3), which overlaps with the −35 hexamer of promoter A2b.

The fact that protein p4 and RNAP bind cooperatively to the promoter suggested that the repression mechanism might involve interaction between the two proteins. Indeed, mutation of protein p4 residue Arg120, which prevents the contact between the two proteins, leads to a loss of repression. As described above, p4 residue Arg120 is also critical for the activation of the late A3 promoter. In this case, Arg120 participates in the stabilization of RNAP at the promoter. At promoter A2c, the result of such interaction is that protein p4 holds the RNAP at the promoter as an initiated complex. Therefore, this region of protein p4 behaves as an activation surface at promoter A3 and as a repression surface at promoter A2c.

Regarding the interaction with the RNAP, protein p4 could form a complex at the A2c promoter with the wild-type α-subunit but not with a deletion mutant lacking the 15 C-terminal amino acids. In addition, promoter repression was impaired when a reconstituted RNAP lacking the 15 C-terminal amino acids of the α-subunit was used. Protein p4 could not interact with this mutant RNAP at promoter A2c (104). Thus, the C-terminal domain of the α-subunit can receive regulatory signals both from transcriptional activators and repressors. On the other hand, the contact between protein p4 and RNAP through the p4 domain containing Arg120 can activate or repress transcription depending on the promoter. Although the interaction with the RNAP occurs through the same surface of protein p4 at the A3 and A2c promoters, the proteins are located at different relative distances in each case: the p4 binding site is centered at position −82 in promoter A3, and at position −71 in promoter A2c. In addition, the A2c promoter contains a good consensus sequence for σ^ARNAP at the −35 hexamer, whereas promoter A3 does not contain the −35 box. Analysis of mutant promoters in which either the distance between the p4 binding site and the transcription start site, or the presence or absence of a −35 box was changed, indicated that the position of protein p4 relative to that of RNAP does not dictate the outcome of the interaction. Rather, it is the absence or presence of a −35 hexamer that determines whether activation or repression occurs. Thus, overstabilization of the RNAP by the presence of a consensus −35 hexamer leads to repression by protein p4, whereas the lack of such a sequence gives rise to p4 activation (101).

The viral double-stranded DNA binding protein p6, involved in repression of the early promoter C2, also plays a role both in the repression of promoters A2c and A2b and in the activation of promoter A3, but only the presence of protein p4 (45). The two proteins, p4 and p6, cooperate with each other in the binding to the central region of the ϕ29 genome containing the A2c, A2b, and A3 promoters, resulting in a ternary p4–p6–DNA complex that affects local DNA topology. Through this complex, protein p6 exerts a role in the repression of promoter A2c impeding unwinding of the DNA strands needed for open complex formation. In contrast, protein p6 functions by reinforcing the positioning of protein p4 in the repression of promoter A2b and activation of promoter A3, thereby facilitating transcription regulation (32). It is interesting to point out that a protein p4 mutant at Arg120, which is not able to interact with the RNAP α-subunit, is able to form the p4–p6–DNA complex and, thus, to both repress the A2c promoter in the presence of protein p6 and to activate the A3 promoter, although in this case to a lower extent than that obtained with the wild-type p4. Other p4 mutants that are unable to form the p4–p6–DNA complex, are also inactive in both activation and repression. Thus, formation of the ternary p4–p6–DNA complex is a requisite of transcription activation and repression mediated by both proteins, p4 and p6 (29).

Taking into account all the results, the switch from early to late transcription at the central promoter region containing the early promoters A2b and A2c and the late promoter A3 can be envisioned as follows. Early after infection, when proteins p4 and p6 have not yet been synthesized in sufficient amounts, the host RNAP is able to transcribe from the early promoters A2b and A2c, but not from the late promoter A3. When both proteins p4 and p6 have been synthesized in sufficient amounts, either of two mechanisms can operate. When only protein p4 binds to a DNA molecule, it stabilizes the binding of RNAP to the late A3 promoter leading to transcription activation. Binding of p4 to its site at the late promoter excludes the RNAP from the early A2b promoter, leading to transcription repression of this promoter. On the other hand, binding of protein p4 to site 2 at the early A2c promoter cooperates with the binding of RNAP giving rise to an overstabilization of the latter, which is unable to leave the promoter, leading also to transcription repression. When both proteins p4 and p6 bind to the same DNA molecule, repression of promoter A2c takes place, because the formation of the open complex is prevented. On the other hand, protein p6 reinforces the positioning of protein p4 at the site located upstream from the A3 promoter (site 3), cooperating in the repression of promoter A2b and in the activation of promoter A3 (32).

Transcription Termination

In phage ϕ29, transcription starting at the late A3 promoter and at the early C2 and C1 promoters terminates in a short intergenic region between gene 16 and ORF 16.5 (see figure 22-1; 4). This DNA region contains an inverted repeat and stem-loop structures with calculated free energies of −14.8 and −16.8 kcal for the early and late transcripts, respectively. In both directions a uridine-rich tail follows the stem-loop, indicating that it functions as a Rho-independent transcriptional terminator, which was named TD1. Similar sequences are present at the corresponding positions in phages B103 and GA-1.

Another Rho-independent transcriptional terminator, named TA1, was found within gene 4 of ϕ29 (4). Thus, some transcripts initiated at promoters A2b and A2c could terminate at TA1, and other transcripts would terminate at the DNA end. This would result in the synthesis of higher levels of messenger RNAs coding for proteins p6 (DBP) and p5 (SSB) than that of messenger RNAs coding for proteins p6 to p1, and could explain the high level of synthesis of proteins p6 and p5 relative to proteins p4 to p1 (reviewed in 90). Similar TA1 transcriptional terminators are present in the B103 and GA-1 DNAs.

In addition, three potential Rho-independent transcriptional terminators are present at the left part of GA-1 DNA (90). The function, if any, of these potential transcriptional terminators remains to be elucidated.

Replication of ϕ29 and its Relatives

Replicative Intermediates

Two types of replicative intermediates have been observed by electron microscopy in ϕ29-infected *B. subtilis*: type I molecules, which are double-stranded DNA with single-stranded tails coming from one or from the two DNA ends, and type II molecules, which are partially double-stranded and partially single-stranded (73, 83). Analysis of these replicative intermediates showed that replication starts at either DNA end, nonsimultaneously, and proceeds towards the other end by strand displacement. However, fully displaced single-stranded DNA is never found. This, as well as other in vivo and in vitro evidence indicates that, before replication started at one DNA end reaches the other end, replication starts at the latter end, and separation of the two displacement forks, when they meet, produces two type II molecules.

Proteins Involved in Replication

By using *ts* and *sus* mutants of phage ϕ29 (93, 105, 139), six genes, 1, 2, 3, 5, 6, and 17, were shown to be involved in the viral DNA replication (36, 72, 114, 140), although genes 1 and 17 were partially dispensable, depending on the growth conditions (25, 39; see later). More recently, gene 16.7 has been characterized, and a supressor-sensitive mutant was constructed. In the absence of p16.7, ϕ29 DNA replication was delayed, suggesting the involvement of this protein in the replication of the viral DNA (92). Genes 2, 3, 5, 6, 16.7, and 17 are also present in phages B103 and GA-1, whereas gene 1 is only present in phage ϕ29.

In vivo ϕ29 DNA replication does not require the *B. subtilis* DNA polymerases I or III (113,140). So far, no bacterial gene has been shown to be required for ϕ29 DNA replication (119).

Protein-Primed Initiation of DNA Replication

It is well known that DNA polymerases are not able to initiate DNA synthesis on a DNA template, but require a primer containing a free hydroxyl (OH) group to start DNA elongation (86). In many cases, RNA primers provide the 3′-OH group needed by the DNA polymerase to elongate the DNA chain. In other cases, the 3′-OH group is created in the DNA template, either by the introduction of a specific nick in one of the strands of a circular double-stranded DNA, or by the formation of hairpin structures at the DNA ends. In the case of most linear DNA genomes containing a protein covalently linked to their 5′ ends (terminal protein, TP), the OH group of a specific serine, threonine or tyrosine residue of the TP is used by the DNA polymerase for initiation (reviewed in 120 and 123). In phage ϕ29, the initiation of replication occurs on the TP-DNA template by deoxynucleotidylation of the TP by the DNA polymerase, resulting in the covalent linkage of 5′-dAMP through a phosphoester bond to the hydroxyl group of Ser232 of the TP (14, 75). This reaction requires the formation of a functional heterodimer between the TP and the DNA polymerase (13). The DNA polymerase active site used for polymerization seems to be used also for the TP deoxynucleotidylation reaction (reviewed in 18), implying a specific positioning of the TP in the heterodimer to allow the DNA polymerase to carry out the reaction. Several mutations in different regions of the DNA polymerase, as well as the absence of the N-terminal domain, have been shown to affect its interaction with the TP (23, 40, 44, 99, 144, 146).

The ϕ29-related phages M2 (89) and Nf (71) belonging to group II, and group III phage GA-1 (80), have been shown to initiate replication by a similar protein-primed mechanism.

In addition, the *Streptococcus pneumoniae* phage Cp-1, the *Escherichia coli* phage PRD1, and adenovirus, also replicate by a protein-priming mechanism (reviewed in 120, 123).

Sliding-Back Mechanism

As already indicated, the DNA ends of ϕ29 have an inverted terminal repeat of six nucleotides (3′-TTTCAT). Interestingly, the linkage of the first dAMP to the TP is not directed by the 3′-terminal nucleotide, but by the second nucleotide from the 3′-DNA end. Then, the TP–dAMP complex slides-back one nucleotide, and the second 3′-terminal nucleotide acts again as template to direct the incorporation of the second dAMP residue (98). For this sliding-back mechanism to take place, a reiteration of at least two nucleotides is required. All the ϕ29-related phages have a reiteration of three Ts at the

3′ ends of their DNAs. Indeed, it has been shown that phage GA-1 also initiates replication at the second nucleotide from the 3′-DNA end (80). Since the TP–dNMP covalent complex is not a substrate of the 3′–5′exonuclease proofreading activity of the DNA polymerase (48) the sliding-back mechanism would be a way to ensure that the initiation of replication occurs with high fidelity.

A terminal reiteration is also present in other DNAs that initiate replication by protein-priming. Thus, the initiation site of replication is also an internal nucleotide close to the 3′-DNA end in the case of phages PRD1 (28) and Cp-1 (87), and in adenovirus (85).

Transition from Protein-Primed to DNA-Primed Replication

The ϕ29 DNA polymerase and the primer TP do not dissociate immediately after initiation, nor after the sliding-back step. There is a stage in which the DNA polymerase synthesizes a DNA molecule of five nucleotides while complexed with the primer TP (initiation mode). Then, during the incorporation of nucleotides 6 to 9 the complex undergoes some structural change (transition mode), and finally the DNA polymerase dissociates from the primer TP when nucleotide 10 is inserted into the nascent DNA chain (elongation mode) (97). These facts probably reflect the requirement of a DNA primer of a minimal length by the DNA polymerase to carry out DNA elongation efficiently. Interestingly, the ϕ29 DNA polymerase mutant, in which Asp456 of the conserved motif YxDTDS (see later) has been changed into Gly, is unable to proceed further than five nucleotides. Thus, residue Asp456 is important to enter into the transition stage of ϕ29 DNA replication (125). A similar transition step takes place in adenovirus replication (84), and is probably a general feature of protein-primed DNA replication.

Replication Origins

The origins of replication of ϕ29 and related phages are the DNA ends with the corresponding TP (parental TP) covalently linked to them. The first step in the initiation of DNA replication is the recognition of the origins by the specific TP/DNA polymerase (pol) heterodimer. Thus, ϕ29 TP-DNA is recognized by the ϕ29 TP/pol heterodimer but not by the Nf TP/pol one, and Nf TP-DNA is better recognized by the Nf TP/pol heterodimer than by the ϕ29 TP/pol one. Nonetheless, ϕ29 TP-DNA can be recognized by the hybrid heterodimer formed by the Nf TP and ϕ29 pol, but not by the one containing the ϕ29 TP and Nf pol. Likewise, Nf TP-DNA is recognized by the hybrid formed by ϕ29 TP and Nf pol but not by the one containing Nf TP and ϕ29 pol. Thus, initiation of replication can occur when TP-DNA and DNA polymerase are from the same phage, indicating that specific recognition of origins is brought about through interaction between DNA polymerase and parental TP (59).

On the other hand, the ϕ29 TP residues Asn80 and Tyr82, which are conserved in the TPs of ϕ29-related phages, have been shown to be involved in the recruitment of the TP/pol heterodimer through interactions between the parental and primer TP (81). These residues are located before a region, spanning amino acids 84 to 118, that has a high probability of forming an amphipathic α-helix, conserved in the TPs of ϕ29-related phages. Indeed, this region of ϕ29 TP has been shown to be an important element for origin recognition (127). Thus, both primer TP and DNA polymerase are involved in interactions with the parental TP for origin recognition.

Blunt-ended DNA fragments containing the left or right ϕ29 DNA ends are active templates for the in vitro initiation of replication, although the activity is greatly reduced with respect to that of TP-containing DNA (53, 70). The terminal 12 bp at each ϕ29 DNA end were shown to be the minimal replication origin (70). On the other hand, efficient initiation requires only the terminal repetition 3′-TT. The 3′-terminal T, although not used as template, increases the affinity of DNA polymerase for the initiation nucleotide (60). Although the parental TP is partially dispensable in vitro, no replication is obtained in vivo after transfection of *B. subtilis* protoplasts with ϕ29 DNA molecules lacking one of the two parental TPs, in agreement with a key role of the parental TP in ϕ29 DNA replication (46).

Essential Proteins Required for ϕ29 DNA Replication and In Vitro Amplification

Genes 2, 3, 5, and 6 of ϕ29 are indispensable for in vivo phage DNA replication. These four genes are conserved in phages B103 and GA-1. As already mentioned, gene 2 codes for the DNA polymerase, gene 3 encodes the TP, gene 5 encodes the SSB protein, and gene 6 codes for the DBP.

An in vitro ϕ29 DNA replication system has been obtained using these four purified proteins. Thus, by using appropriate amounts of purified TP, DNA polymerase, DBP, and SSB, it has been possible to amplify in vitro small amounts (0.5 ng) of ϕ29 TP-DNA (19,285 bp long) by 3 orders of magnitude (0.5 μg) after 1 hour of incubation at 30 °C. The fidelity of the amplified DNA was demonstrated since its infectivity, measured as the ability to produce phage particles in transfection experiments, was identical to that of the natural ϕ29 DNA obtained from virions (12).

Primer TP

As already mentioned, the primer TP forms an heterodimer with the DNA polymerse for the initiation of TP-primed DNA replication. Several lines of evidence indicate that the TP occupies the double-stranded DNA binding channel in the DNA polymerase during initiation of replication (40, 144). Heterodimer formation involves several contacts with different regions of the ϕ29 TP, as it was indicated before to be the case for the ϕ29 DNA polymerase. These include an internal region spanning amino acids 72 to 80, amino acids 242 to 262 at the C-end (154), and the RGD motif located at positions 256 to 258 in the ϕ29 TP (82). The RGD motif is conserved in most ϕ29-related phages, but not in phage GA-1.

In ϕ29, the 5′-terminal dAMP is linked through a phosphoester bond to the hydroxyl group of Ser232 of the TP (75). This amino acid residue is highly specific since a mutant ϕ29 TP, in which Ser232 was replaced by Thr, completely lost its priming activity (55). Residue Ser232 is conserved in the TP of phages B103 and GA-1, suggesting that it is also used for the covalent linkage of the first dAMP in both phages.

DNA Polymerase

DNA polymerase is encoded by gene 2 in phages ϕ29, B103, and GA-1. In ϕ29, Nf (belonging to group II), and GA-1, the DNA polymerase has been shown to be required for the replication of the viral DNA (14, 59, 80, 150). The ϕ29 DNA polymerase is inhibited by aphidicolin, phosphonoacetic acid, and the nucleotide analogs butylanilino-dATP and butylphenyl-dGTP, known inhibitors of eukaryotic DNA polymerase α. These functional, as well as structural criteria have allowed the classification of the ϕ29 DNA polymerase as belonging to the B-type superfamily of DNA-dependent DNA polymerases (also named eukaryotic- or α-like DNA polymerases). ϕ29 DNA polymerase, which is a monomer of about 65 kDa, catalyzes both the initiation and elongation stages of DNA synthesis (14, 15). Thus, it is able to carry out two different reactions: TP-deoxynucleotidylation and DNA polymerization. In addition, it has two degradative activities: pyrophosphorolysis (19) and 3′–5′ exonucleolysis (16, 151). Moreover, it has two very important intrinsic properties: high processivity and strand displacement activity (10). The insertion discrimination values of the ϕ29 DNA polymerase range from 10^4 to 10^6, and the efficiency of mismatch elongation is 10^5- to 10^6-fold lower than that of a paired terminus (48). This high fidelity of ϕ29 DNA polymerase is due to its 3′–5′ exonuclease activity (54).

Site-directed and deletion mutagenesis have shown that the ϕ29 DNA polymerase has a bimodular organization. The N-terminal domain contains the 3′–5′ exonuclease activity, as well as the strand displacement ability, and the C-terminal domain contains the 5′–3′ synthetic activities, initiation and polymerization, as well as the pyrophosphorolytic activity (reviewed in 17, 18).

N-Terminal Domain

Three conserved regions, named ExoI, ExoII, and ExoIII, form the 3′–5′ exonuclease active site and are evolutionarily conserved in prokaryotic and eukaryotic DNA polymerases (6). The three Exo motifs contain five invariant residues involved in metal binding and 3′–5′ exonuclease catalysis that, in the case of ϕ29 DNA polymerase, are Asp12 and Glu14 in ExoI, Asp66 in ExoII, and Tyr165 and Asp169 in ExoIII (6). Interestingly, mutations in these five amino acid residues, which strongly reduced the 3′–5′ exonuclease activity, were also greatly affected in their strand displacement capacity (49, 135). Another invariant residue, Lys143 of ϕ29 DNA polymerase, was shown to be important both for the catalytic efficiency of the 3′–5′ exonuclease activity and in its strand displacement capacity (41).

Other residues in the Exo motifs, conserved in most prokaryotic and eukaryotic DNA polymerases, as well as in those of phages B103 and GA-1, were also analyzed by site-directed mutagenesis. Two of these residues, Thr15 and Asn62, located at the ExoI and ExoII motifs, respectively, act as single-stranded DNA ligands, having a critical role in the stabilization of the frayed primer terminus at the 3′–5′ exonuclease active site, but they do not affect the strand displacement activity (42). In addition, Phe65 of the Exo II motif, and residues Ser122 and Leu123, which are part of a newly identified motif between motifs ExoII and ExoIII, act as single-stranded DNA ligands for 3′–5′ exonucleolysis (43).

C-Terminal Domain

This region of the ϕ29 DNA polymerase contains five motifs also conserved in other DNA polymerases that belong to family B. These motifs are Dx_2SLYP (motif A or 1), Kx_3NSxYG (motif B or 2a), YxDTDS (motif C or 3), Tx_2GR (motif 2b), and KxY (motif 4). Mutational analysis showed that this domain of the ϕ29 DNA polymerase contains the polymerization and protein-primed initiation activity, containing sites for interaction with the metal activator, dNTPs, TP, and DNA. For several DNA polymerases, including the ϕ29 DNA polymerase, three Asp residues are involved in metal binding and catalysis at the polymerization active site. These three residues are Asp249 (motif Dx_2SLYP), Asp456, and Asp458 (motif YxDTDS) in the ϕ29 DNA polymerase (7, 21). In addition, ϕ29 DNA polymerase residue Arg438 (motif Tx_2GR) plays a role in catalysis of the polymerization reaction (99). Three Tyr residues, highly conserved in family B DNA polymerases, have been shown to be involved in interaction with dNTPs. In the ϕ29 DNA polymerase these residues are Tyr254 (motif Dx_2SLYP; 20), Tyr390 (motif Kx_3NSxYG; 20, 22), and Tyr454 (motif YxDTDS; 7). Residues Tyr254 and Tyr390 are also involved in nucleotide binding selection, thus playing an important role in the fidelity of DNA replication (124). In addition, a specific change of Tyr254 into Val enables the mutant ϕ29 DNA polymerase to incorporate ribonucleotides without affecting its capacity to incorporate dNTPs (24). This indicates that Tyr254 of ϕ29 DNA polymerase is responsible for the discrimination against the 2′-OH group of an incoming ribonucleotide. Another ϕ29 DNA polymerase residue, Lys383, invariant in motif Kx_3NSxYG of family B DNA polymerases, is also involved in dNTP binding (126).

On the other hand, eight residues highly conserved in family B DNA polymerases were shown to be involved in binding template-primer structures in the ϕ29 DNA polymerase. These residues are Ser252 (motif Dx_2SLYP; 21), Asn387, Gly391, and Phe93 (motif Kx_3NSxYG; 22), Thr434 and Arg438 (motif Tx_2GR; 99), and Lys498 and Tyr500 (motif KxY; 23).

Coordination Between Synthesis and Degradation

A conserved motif, YxGG/A, located between the exonuclease and polymerization domains of family B DNA polymerases, is a DNA binding motif that plays an important role in the coordination between DNA synthesis and proofreading (145). Moreover, this motif has been shown to be important for the formation of a stable complex between TP and ϕ29 DNA polymerase, affecting initiation and transition in ϕ29 TP-DNA replication (144).

A C-terminal deletion derivative of ϕ29 DNA polymerase, containing the first 188 N-terminal amino acids, was devoid of synthetic activities (TP-primed initiation and DNA polymerization), but retained some 3′–5′ exonuclease activity (17). In addition, an N-terminal deletion derivative, containing amino acids 189 through 575, lacked 3′–5′ exonuclease activity and strand displacement capacity. It retained some polymerization activity, although this was decreased with respect to that of the complete polymerase, being distributive instead of processive. These polymerization defects could be related to a strong impairment in DNA binding, suggesting that contacts present in the N-terminal domain are important for an optimal stabilization and translocation of the DNA during polymerization. In addition, the C-terminal domain showed a reduced initiation of TP-primed DNA replication due to a reduced capacity to interact with the primer TP and a lack of activation by protein p6 (146).

Protein p6 (DBP)

The double-stranded DNA binding protein p6 (DBP), described as a histone-like protein, is able to bind in vitro to the whole ϕ29 DNA, and a role in genome organization has been proposed (68, 129). Protein p6 binds preferentially to the ϕ29 DNA ends (115, 131) through the minor groove (50). The main p6 binding sites are located at nucleotides 46 to 68 and 62 to 125 at the left and right ϕ29 DNA ends, respectively. These regions do not have sequence similarity, but they contain sequences that are predicted to be bendable every 12 bp (131), this feature being the main determinant for p6 recognition (130).

In addition to the role of p6 in transcription regulation (see above), its binding to the ϕ29 DNA ends activates the initiation of DNA replication (11). When protein p6 binds to circular DNA it restrains positive supercoiling, supporting a model in which a right-handed DNA superhelix wraps tightly around a multimeric protein p6 core (132). The DNA in the complex with p6 is compacted 4.2-fold. The parameters that define the path followed by the DNA in the p6 complex have been determined: one superhelical turn has 63 bp with a pitch of 5.1 nm and a diameter of 6.6 nm. Thus, the DNA should be bent (66° every 12 bp) and underwound (11.5 bp per turn). It is believed that the conformation of the nucleoprotein complex would help to open the DNA ends facilitating the formation of the TP-dAMP initiation complex (122).

By site-directed and deletion mutagenesis it has also been shown that the N-terminal region of p6 is involved in DNA binding. Specifically, mutations in residues Lys2 and Arg6 produced p6 proteins affected in DNA binding (50).

Protein p6 was shown to form dimers in solution (110). Residues critical for the self-association of the protein, identified by random mutagenesis, are Ile8 and Ala44. Mutations at these two residues showed, in addition to impaired dimer formation ability, reduced DNA binding affinity and they were affected in the initiation of DNA replication (1). Thus, dimers seem to be the active form of p6 for DNA binding.

Protein p6 binds with a defined phase to the ϕ29 DNA ends, which is essential for the activation of the initiation of DNA replication (132). By using the protein p6 counterparts of phages Nf (of the same group as B103) and GA-1, as well as the TP and DNA polymerase from these phages, in addition to the TP-DNA complexes of ϕ29, Nf, and GA-1, it has been shown that the activation of the initiation of replication requires not only the formation of a specific nucleoprotein complex but also its specific recognition by the proteins involved in the initiation of DNA replication (51).

Protein p5 (SSB)

The ϕ29 SSB protein, product of gene 5, is essential for the elongation of in vivo viral DNA replication (94). Binding of ϕ29 SSB to the single-stranded DNA produced during ϕ29 DNA replication in vitro has been demonstrated (69). This binding stimulates dNTP incorporation during ϕ29 DNA replication (88), and increases the elongation rate, mainly when ϕ29 DNA polymerase mutants impaired in strand displacement are used (135). This effect is probably due to the helix destabilizing activity of the SSB protein (137). When ϕ29 DNA amplification assays are carried out in the absence of SSB, short ϕ29 DNA products are produced that have a palindromic nature and are caused by DNA polymerase template-switching (12, 47).

ϕ29 SSB binds single-stranded DNA in a cooperative way ($K_{\mathrm{eff}} = 10^5\ \mathrm{M}^{-1}$; $w = 50$–70), covering 3.4 nucleotides per monomer (136). A comparative study of the structural complexes formed by the ϕ29, Nf, and GA-1 SSB with single-stranded DNA has been carried out (57). The SSB of ϕ29 and Nf are monomers in solution, whereas GA-1 SSB behaves as a hexamer. The binding site size of the latter is 51 nucleotides per hexamer, compared with 3.4 and 4.7 nucleotides per monomer, respectively, for ϕ29 and Nf SSB. In addition, the length of circular single-stranded DNA was reduced about 6-fold upon GA-1 SSB binding, and only 2-fold when ϕ29 SSB was bound to the DNA. In agreement with the structural features, less GA-1 SSB than ϕ29 or Nf SSB was needed for a similar helix destabilizing activity and to stimulate in vitro DNA replication (58). Comparison of the amino acid sequence of the three SSB showed that GA-1 SSB has an N-terminal addition of about 40 amino acids. Deletion analysis has indicated that the region comprising amino acids 19 to 26 is essential for GA-1 SSB oligomerization (56).

Other Proteins Involved in DNA Replication

Protein p1

Phage ϕ29 DNA replication was strongly reduced when nonsuppressor *B. subtilis* cells were infected with mutant phage *sus*1 (629) at 37 °C (25, 114). However, infection at 30 °C reduced phage DNA synthesis to a lesser extent (26). Protein p1 associates with the cell membrane, the 43 C-terminal amino acids being required for this association (25). In addition, protein p1 lacking the 33 N-terminal amino acids assembled into long protofilaments associated in a highly ordered, parallel array forming two-dimensional sheets (26). Protein p1 has been shown to interact with the TP in vitro (27). Altogether, the results suggest that protein p1 is a component of a virus-associated membrane structure which would provide an anchoring site for the phage DNA replication machinery.

As already indicated, phage B103 and GA-1 do not contain a corresponding gene 1. However, the deduced protein sequence of ORF h of phage B103 shows significant similarity (62.5%) to the 46 C-terminal amino acids of ϕ29 protein p1. The function, if any, of ORF h of phage B103 remains to be determined.

Protein p17

When nonsuppressor *B. subtilis* cells were infected with the ϕ29 mutant *sus*17 (112), viral DNA synthesis was reduced (36). In addition, infection with the *sus*17 (112) mutant in solid media produced very tiny plaques (93). More recently, it has been shown that viral DNA synthesis after *sus*17 (112) infection depends on the multiplicity of infection. At a low multiplicity, viral DNA synthesis is strongly reduced, whereas a high multiplicity of infection results in a wild-type phenotype (39). In addition, a stimulatory effect by protein p17 of in vitro ϕ29 DNA amplification was observed under conditions of limiting amounts of DNA and initiation proteins (39). It has been shown that protein p17 interacts in vitro with protein p6 and stimulates binding of the latter to the ϕ29 DNA ends (38).

The gene 17 corresponding to other ϕ29-related phages is organized differently, with deletions at several regions, even in the case of phages that belong to group I, such as PZA and ϕ15 (5, 109). A comparison of gene 17 of different ϕ29-related phages has been carried out (111).

Protein p16.7

ORF 16.7, located at the right end of the ϕ29 genome, encodes a protein of 130 amino acids, p16.7, which is expressed early after infection. Computer-assisted analysis showed that the 22 N-terminal amino acids of p16.7 have a very hydrophobic character and may constitute a transmembrane-spanning domain. On the other hand, the region spanning amino acids 19 to 60 has a high potential to form an α-helical coiled-coil structure and, thus, could function as an oligomerization domain. In addition, the C-terminal part of p16.7 (amino acids 70 to 130) shows some similarity to DNA binding proteins. Indeed, it has been shown that p16.7 is a membrane protein, and the N-terminal transmembrane domain is required for its membrane localization (92). A variant protein, p16.7A, was constructed, in which the N-terminal membrane anchor domain was replaced by a histidine tag. The purified protein was shown to form dimers in solution and to bind to single-stranded DNA. In fact, it binds in vitro to the displaced strands of ϕ29 replicative intermediates, and it is able to replace the ϕ29 SSB protein in ϕ29 DNA amplification assays (128). However, unlike the ϕ29 SSB, protein p16.7 does not have helix-destabilizing activity.

To study the in vivo role of protein p16.7, a ϕ29 mutant containing a suppressible mutation in gene 16.7 was constructed. *In vivo* phage DNA replication was significantly delayed after restrictive infection with the *sus*16.7 mutant (92). It was also found that redistribution of replicating phage DNA from the initial replication site to various sites surrounding the nucleoid was also dependent on protein p16.7 (91). A model has been proposed in which the main role of p16.7 in vivo involves the attachment of replicating ϕ29 DNA molecules to the membrane of infected cells (128). Homologs of gene 16.7 are present in phages B103 and GA-1, suggesting that the proposed role of p16.7 is conserved in this family of phages.

Phage Morphogenesis

Phage morphogenesis has been thoroughly studied in ϕ29 (for a review see 2), and a very efficient in vitro system for DNA packaging has been established (8, 64). In addition, the three-dimensional structure of the ϕ29 particle and that of the empty proheads have been determined (141). Both infection with ϕ29 mutants and in vitro assembly have been used to establish the morphogenesis pathway of ϕ29. No similar studies have been carried out in phages B103 or GA-1.

Prohead Formation

The proheads of ϕ29, which have the form of a prolate icosahedron, contain 235 copies of the major capsid protein (p8), about 180 copies of the scaffolding protein (p7), 55 dimers of the head fiber protein (p8.5), 12 copies of the head–tail connector (p10), six copies of the 174-nucleotide-long pRNA, and five or six copies of the ATPase (p16). The major capsid proteins are hexameric structures located at each of the 3-fold axes and pentameric structures at each of the 5-fold axes. The head fibers attach as dimers to the p8 subunits at quasi-3-fold axes that relate one pentamer to a pair of hexamers (141). Isometric particles are formed in gene 10 (coding for the connector) mutant-infected cells (33), suggesting that the connector is the structure from which capsid assembly is initiated (63).

DNA Packaging Machine

The head–tail connector is an oligomer with 12-fold symmetry (35) whose structure and topology have been extensively studied (62, 79, 106, 133, 147, 148). The connector structure can be divided in three regions: a narrow end, a central part, and a wide end. The wide end of the connector, which has 12-fold symmetry, is buried inside the prohead, whereas the narrow end, with an apparent 6-fold symmetry, protrudes from the portal vertex of the prohead.

Packaging of ϕ29 DNA into the prohead, both in vivo and in vitro, requires a specific small RNA (pRNA) encoded at the left of the genome (65, 143). The pRNA forms a hexameric ring structure (67, 76, 155) with a diameter similar to that of the narrow end of the connector, which results in the superposition of the pRNA hexamer on the connector, forming a double-ring structure (155). This complex is essential for DNA packaging; it recognizes ϕ29 DNA, and it is probably responsible for the specificity of packaging from the left DNA end (9).

Protein p16 of ϕ29 has ATPase activity dependent on DNA, on proheads, and on pRNA (61, 66). Five or six copies of protein p16 are required for ϕ29 DNA packaging (133). Thus, the structure formed by 12 molecules of the connector and six pRNA molecules forms a very efficient DNA-translocating motor that, with the aid of the DNA packaging protein p16 and ATP, actively pumps the ϕ29 DNA into the prohead (for a review see 74 and chapter 6). This motor can work against loads up to 57 pN on average, making it one of the strongest molecular motors reported to date (134). Phage ϕ29 DNA packaging is a very energy-consuming process requiring 1 ATP molecule to package 2 bp of DNA (66).

Phage Maturation

During DNA packaging, the prohead becomes more angular and rigid, and the scaffolding protein p7 is released (141). After packaging, the ATPase protein p16 and the pRNA molecules are also released from the prohead (66). Then, six copies of the lower collar protein (p11), three or four copies of the tail protein (p9), and 12 copies of the appendages (dimers of protein p12* cleaved from p12 precursor molecules) are assembled sequentially (34, 52; for a review see 2). For a stable assembly of the tail protein, the nonstructural protein p13 is required (52). The lower collar has, like the distal end of the connector, a 6-fold symmetry, which may explain the high stability of the neck complex (37). Removal of the tail protein results in the release of the DNA from the particles, indicating that it functions as a stop for DNA exit (34).

Phage Lysis

In the three phages ϕ29, B103, and GA-1, genes 14 and 15 encode a holin and a peptidoglycan hydrolase, respectively, both of which are required for efficient lysis of the infected bacteria (see chapter 10). Lysis is delayed in cells infected with a ϕ29 *sus*14 mutant, giving rise to a larger burst size than after infection with wild-type phage (36). Holins are membrane proteins that introduce pores in the cell membrane, allowing the peptidoglycan hydrolase to exit the cytoplasm and hydrolyze the cell wall. The ϕ29 holin protein (p14) contains two or three potential transmembrane domains (138). It also contains two start codons at positions 1 and 3, each with a ribosomal binding site, giving rise to two products of 131 and 129 amino acids respectively (142). The two proteins have opposite functions, the longer product acting as an inhibitor of the shorter one, the lysis effector. Cooperative action of the inhibitor and effector results in the proper timing of cell lysis. Gene 14 of phage B103 also contains the dual start motif, whereas gene 14 of GA-1 only has the start codon that would give rise to the lysis effector.

The peptidoglycan hydrolases encoded by phages ϕ29 and B103, also named lysozymes, belong to the group of muramidases (112, 118). The peptidoglycan hydrolase encoded by gene 15 of phage GA-1 only has moderate homology with those of ϕ29 and B103 (see table 22-1; 90). In fact, the gene 15 product of phage GA-1 is more closely related to several *Bacillus*-encoded autolysins belonging to the group of amylases.

Conclusions

The molecular mechanisms underlying the different stages in the development of phage ϕ29 have been unraveled through more than 30 years of research. In several respects, ϕ29 turns out to be a model system in studies of replication, regulation of transcription, and phage morphogenesis. In replication, the novel protein-primed mechanism has been worked out, with the finding of a unique DNA polymerase that, by itself, is highly processive and able to produce strand displacement. Using the four main replication proteins, in vitro ϕ29 DNA amplification has been obtained. In transcriptional regulation, the concerted action of two proteins can operate. One is a sequence-specific regulatory protein that acts as an activator or as a repressor, depending upon the context of the promoter. The other is a histone-like protein that cooperates with the regulatory protein, both in activation and in repression. In phage morphogenesis, a very efficient system of DNA packaging in vitro is available, with the interesting finding of a small RNA molecule which is required in the packaging event. Although much is known regarding the major mechanisms used in phage development, there are still some minor aspects that remain to be known: for example, the elucidation of the role of several ORFs in phage ϕ29 and B103, as well as that of the additional information present in phage GA-1 that is lacking in phages ϕ29 and B103.

Acknowledgments

I am grateful to Dr. W. Meijer for the preparation of figure 22-1. This work was supported by research grants 2R01 GM27242–22 from the National Institutes of Health, PB98–045 from Dirección General de Investigación Científica y Técnica, and an Institutional grant from Fundación Ramón Areces.

References

1. Abril, A. M., M. Salas, and J. M. Hermoso. 2000. Identification of residues within two regions involved in self-association of viral histone-like protein p6 from phage ϕ29. J. Biol. Chem. 275:26404–26410.
2. Anderson, D. L., and B. E. Reilly. 1993. pp. 859–867, *In* L. Sonenshein, J. A. Hoich, and R. Losick, (eds.) Morphogenesis of bacteriophage ϕ29, *Bacillus Subtilis* and Other Gram-Positive Bacteria: Biochemistry, Physiology, and Molecular Genetics. American Society for Microbiology, Washington, D.C.
3. Barthelemy, I., R.P Mellado, and M. Salas. 1989. In vitro transcription of bacteriophage ϕ29 DNA: inhibition of early promoters by the viral replication protein p6. J. Virol. 63:460–462.
4. Barthelemy, I., M. Salas, and R. P. Mellado. 1987. In vivo transcription of bacteriophage ϕ29 DNA. Transcription termination. J. Virol. 61:1751–1755.
5. Benes, V., L., Arnold, J., Smrt, and V. Paces. 1989. Nucleotide sequence of the right early region of *Bacillus* phage ϕ15 and comparison with related phages: reorganization of gene 17 during evolution. Gene 75:341–347.
6. Bernad, A., L. Blanco, J. M. Lázaro, G. Martín, and M. Salas. 1989. A conserved 3′ → 5′ exonuclease active site in prokaryotic and eukaryotic DNA polymerases. Cell 59:219–228.
7. Bernad, A., J. M. Lázaro, M. Salas, and L. Blanco. 1990. The highly conserved amino acid sequence motif Tyr-Gly-Asp-Thr-Asp-Ser in a-like DNA polymerases is required by phage ϕ29 DNA polymerase for protein-primed initiation and polymerization. Proc. Natl. Acad. Sci. USA 87:4610–4614.
8. Bjornsti, M. A., B. E. Reilly, and D. L. Anderson. 1981. In vitro assembly of the *Bacillus subtilis* bacteriophage ϕ29. Proc. Natl. Acad. Sci. USA 78:5861–5865.
9. Bjornsti, M. A., B. E. Reilly, and D. L. Anderson. 1983. Morphogenesis of bacteriophage ϕ29 of *Bacillus subtilis*: oriented and quantized in vitro packaging of DNA protein gp3. J. Virol. 45:383–396.
10. Blanco, L., A. Bernad, J. M. Lázaro, G. Martín, C. Garmendia, and M. Salas. 1989. Highly efficient DNA synthesis by the phage ϕ29 DNA polymerase. Symmetrical mode of DNA replication. J. Biol. Chem. 264:8935–8940.
11. Blanco, L., J. Gutiérrez, J. M. Lázaro, A. Bernad, and M. Salas. 1986. Replication of phage ϕ29 DNA in vitro: role of the viral protein p6 in initiation and elongation. Nucleic Acids Res. 14:4923–4937.
12. Blanco, L., J. M. Lázaro, M. de Vega, A. Bonnin, and M. Salas. 1994. Terminal protein-primed DNA amplification. Proc. Natl. Acad. Sci. USA 91:12198–12202.
13. Blanco, L., I. Prieto, J. Gutiérrez, A. Bernad, J. M. Lázaro, J. M. Hermoso, and M. Salas. 1987. Effect of NH4+ ions on ϕ29 DNA-protein p3 replication: formation of a complex between the terminal protein and the DNA polymerase. J. Virol. 61:3983–3991.
14. Blanco, L., and M. Salas. 1984. Characterization and purification of a phage ϕ29 coded DNA polymerase required for the initiation of replication. Proc. Natl. Acad. Sci. USA 81:5325–5329.
15. Blanco, L., and M. Salas. 1985. Replication of ϕ29 DNA with purified terminal protein and DNA polymerase: synthesis of full-length ϕ29 DNA. Proc. Natl. Acad. Sci. USA 82:6404–6408.
16. Blanco, L., and M. Salas. 1985. Characterization of a 3′→ 5′ exonuclease activity in the phage ϕ29-encoded DNA polymerase. Nucleic Acids Res. 13:1239–1249.
17. Blanco, L., and M. Salas. 1995. Mutational analysis of bacteriophage ϕ29 DNA polymerase. Methods Enzymol. 262:283–294.
18. Blanco, L., and M. Salas. 1996. Relating structure to function in ϕ29 DNA polymerase. J. Biol. Chem. 271:8509–8512.
19. Blasco, M. A., A. Bernad, L. Blanco, and M. Salas. 1991. Characterization and mapping of the pyrophosphorolytic activity of the phage ϕ29 DNA polymerase. J. Biol. Chem. 266:7904–7909.
20. Blasco, M. A., J. M. Lázaro, A. Bernad, L. Blanco, and M. Salas. 1992. ϕ29 DNA polymerase active site: mutants in conserved residues Tyr254 and Tyr390 are affected in dNTP binding. J. Biol. Chem. 267:19427–19434.

21. Blasco, M. A., J. M. Lázaro, L. Blanco and M. Salas. 1993. ϕ29 DNA polymerase active site. Residue Asp^{249} of conserved amino acid motif Dx$_2$SLYP is critical for synthetic activities. J. Biol. Chem. 268:24106–24113.
22. Blasco, M. A., J. M. Lázaro, L. Blanco, and M. Salas. 1993. ϕ29 DNA polymerase active site. The conserved amino acid motif "Kx$_3$NsxYG" is involved in template-primer binding and dNTP selection. J. Biol. Chem. 268:16763–16770.
23. Blasco, M. A., J. Méndez, J. M. Lázaro, L. Blanco, and M. Salas. 1995. Primer-terminus stabilizatión at the ϕ29 DNA polymerase active site: mutational analysis of conserved motif KxY. J. Biol. Chem. 270:2735–2740.
24. Bonnin, A., J. M. Lázaro, L. Blanco, and M., Salas. 1999. A single tyrosine prevents insertion of ribonucleotides in the eukaryotic-type ϕ29 DNA polymerase. J. Mol. Biol. 290:241–251.
25. Bravo, A., and M. Salas. 1997. Initiation of bacteriophage ϕ29 DNA replication in vivo: assembly of a membrane-associated multiprotein complex. J. Mol. Biol. 269:102–112.
26. Bravo, A., and M. Salas. 1998. Polymerization of bacteriophage ϕ29 replication protein p1 into protofilament sheets. EMBO J. 17:6096–6105.
27. Bravo, A., B. Illana, and M. Salas. 2000. Compartmentalization of phage ϕ29 DNA replication: interaction between the primer terminal protein and the membrane-associated protein p1. EMBO J. 19:5575–5584.
28. Caldentey, J., L. Blanco, D. H. Bamford, and M. Salas. 1993. In vitro replication of bacteriophage PRD1 DNA. Characterization of the protein-primed initiation site. Nucleic Acids Res. 21:3725–3730.
29. Calles, B., M. Salas, and F. Rojo. 2002. The ϕ29 transcriptional regulator contacts the nucleoid protein p6 to organize a repression complex. EMBO J. 21:6185–6194.
30. Camacho, A., and M. Salas. 1999. Effect of mutations in the "Extended-10" motif of three *Bacillus subtilis* σA-RNA polymerase-dependent promoters. J. Mol. Biol. 286:683–693.
31. Camacho, A., and M. Salas. 2001. Repression of bacteriophage ϕ29 early promoter C2 by the viral protein p6 is due to the impairment of the closed complex. J. Biol. Chem. 276:28927–28932.
32. Camacho, A., and M. Salas. 2001. Mechanism for the switch of ϕ29 DNA early to late transcription by regulatory protein p4 and histone-like protein p6. EMBO J. 20:6060–6070.
33. Camacho, A., F. Jiménez, J. de la Torre, J. L. Carrascosa, R. P. Mellado, C. Vásquez, E. Viñuela, and M. Salas. 1977. Assembly of *B. subtilis* phage ϕ29. I. Mutants in the cistrons coding for the structural protein. Eur. J. Biochem. 73:39–55.
34. Camacho, A., F., Jiménez, E. Viñuela, and M. Salas. 1979. Order of assembly of the lower collar and the tail proteins of *Bacillus subtilis* phage ϕ29. J. Virol. 29:540–545.
35. Carazo, J. M., L. E. Donate, L. Herranz, J. P. Secilla, and J. L. Carrascosa. 1986. Three-dimensional reconstruction of the connector of bacteriophage ϕ29 at 1.8 nm resolution. J. Mol. Biol. 192:853–867.
36. Carrascosa, J. L., A. Camacho, F. Moreno, F. Jiménez, R. P. Mellado, E. Viñuela, and M. Salas. 1976. *Bacillus subtilis* phage ϕ29: characterization of gene products and functions. Eur. J. Biochem. 66:229–241.
37. Carrascosa, J. L., J. M. Carazo, and N. García. 1983. Structural localization of the proteins of the head to tail connecting region of bacteriophage ϕ29. Virology 124:133–144.
38. Crucitti, P., A. M. Abril, and M. Salas. 2003. Bacteriophage ϕ29 early protein p17. Self-association and hetero-association with the viral histone-like protein p6. J. Biol. Chem. 278:4906–4911.
39. Crucitti, P., J. M. Lázaro, V. Benes, and M. Salas. 1998. Bacteriophage ϕ29 early protein p17 is conditionally required for the first rounds of viral DNA replication. Gene 223:135–142.
40. de Vega, M., L. Blanco, and M. Salas. 1998. ϕ29 DNA polymerase residue Ser122, a single-stranded DNA ligand for 3′–5′ exonucleolysis, is required to interact with the terminal protein. J. Biol. Chem. 273:28966–28977.
41. de Vega, M., T. Ilyina, J. M. Lázaro, M. Salas, and L. Blanco. 1997. An invariant lysine residue is involved in catalysis at the 3′–5′ exonuclease active site of eukaryotic-type DNA polymerases. J. Mol. Biol. 270:65–78.
42. de Vega, M., J. M. Lázaro, M. Salas, and L. Blanco. 1996. Primer-terminus stabilization at the 3′–5′ exonuclease active site of ϕ29 DNA polymerase. Involvement of two amino acid residues highly conserved in proofreading DNA polymerases. EMBO J. 15:1182–1192.
43. de Vega, M., J. M. Lázaro, M. Salas, and L. Blanco. 1998. Mutational analysis of ϕ29 DNA polymerase residues acting as single-stranded DNA ligands for 3′–5′ exonucleolysis. J. Mol. Biol. 279:807–822.
44. Dufour, E., J. Méndez, J. M. Lázaro, M. de Vega, L. Blanco, and M. Salas. 2000. An aspartic acid residue in TPR-1, a specific region of protein-priming DNA polymerases, is required for the functional interaction with primer terminal protein. J. Mol. Biol. 304:289–300.
45. Elías-Arnanz, M., and M. Salas. 1999. Functional interactions between a phage histone-like protein and a transcriptional factor in regulation of ϕ29 early–late transcriptional switch. Genes Dev. 13:2502–2513.
46. Escarmís, C., D. Guirao, and M. Salas. 1989. Replication of recombinant ϕ29 DNA molecules in *Bacillus subtilis* protoplasts. Virology 169:152–160.
47. Esteban, J. A., L. Blanco, L. Villar, and M. Salas. 1997. In vitro evolution of terminal protein-containing genomes. Proc. Natl. Acad. Sci. USA 94:2921–2926.
48. Esteban, J. A., M. Salas, and L. Blanco. 1993. Fidelity of ϕ29 DNA polymerase. Comparison between protein-primed initiation and DNA polymerization. J. Biol. Chem. 268:2719–2726.
49. Esteban, J. A., M. S. Soengas, M. Salas and L. Blanco. 1994. 3′ → 5′ exonuclease active site of ϕ29 DNA polymerase. Evidence favoring a metal-ion-assisted reaction mechanism. J. Biol. Chem. 269:31946–31954.
50. Freire, R., M. Salas, and J. M. Hermoso. 1994. A new protein domain for binding to DNA through the minor groove. EMBO J. 13:4353–4360.

51. Freire, R., M. Serrano, M. Salas, and J. M. Hermoso. 1996. Activation of replication origins in φ29-related phages requires the recognition of initiation proteins to specific nucleoprotein complexes. J. Biol. Chem. 271:31000–31007.
52. García, J. A., J. L. Carrascosa, and M. Salas. 1983. Assembly of the tail protein of the *Bacillus subtilis* phage φ29. Virology 12:18–30.
53. García, J. A., M. A. Peñalva, L. Blanco, and M. Salas. 1984. Template requirements for initiation of phage φ29 DNA replication in vitro. Proc. Natl. Acad. Sci. USA 81:80–84.
54. Garmendia, C., A. Bernad, J. A. Esteban, L. Blanco, and M. Salas. 1992. The bacteriophage φ29 DNA polymerase, a proofreading enzyme. J. Biol. Chem. 267:2594–2599.
55. Garmendia, C., M. Salas, and J. M. Hermoso. 1988. Site-directed mutagenesis of bacteriophage φ29 terminal protein: isolation and characterization of a Ser232 → thr mutant in the linking site. Nucleic Acids Res. 16:5727–5740.
56. Gascón, I., J. L. Carrascosa, L. Villar, J. M., Lázaro, and M. Salas. 2002. Importance of the N-terminal region of the phage GA-1 SSB for its self-interaction ability and functionality. J. Biol. Chem. 277:22534–22540.
57. Gascón, I., C. Gutiérrez, and M. Salas. 2000. Structural and functional comparative study of the complexes formed by viral φ29, Nf and GA-1 SSB proteins with DNA. J. Mol. Biol. 296:989–999.
58. Gascón, I., J. M. Lázaro, and M. Salas. 2000. Comparative functional characterization of viral φ29, Nf and GA-1 SSB proteins. Nucleic Acids Res. 28:2034–2042.
59. González-Huici, V., J. M. Lázaro, M. Salas, and J. M. Hermoso. 2000. Specific recognition of parental terminal protein by DNA polymerase for initiation of protein-primed DNA replication. J. Biol. Chem. 275:14678–14683.
60. González-Huici, V., M. Salas, and J. M. Hermoso. 2000b. Sequence requirements for protein-primed initiation and elongation of phage φ29 DNA replication. J. Biol. Chem. 275:14678–14683.
61. Grimes, S., and D. Anderson. 1990. RNA dependence of the bacteriophage φ29 DNA packaging ATPase. J. Mol. Biol. 215:559–566.
62. Guasch, A., J. Pous, A. Párraga, J. M. Valpuesta, J. M. Carrascosa, and M. Coll. 1998. Crystallographic analysis reveals the 12-fold symmetry of the bacteriophage φ29 connector particles. J. Mol. Biol. 281:219–225.
63. Guo, P., S. Erickson, W. Xu, N. Olson, T. S. Baker, and D. Anderson. 1991. Regulation of the phage φ29 prohead shape and size by the portal vertex. Virology 183:366–373.
64. Guo, P., S. Grimes, and D. Anderson. 1986. A defined system for in vitro packaging of DNA-gp3 of the *Bacillus subtilis* bacteriophages φ29. Proc. Natl. Acad. Sci. USA 83:3505–3509.
65. Guo, P., C. Peterson, and D. Anderson. 1987. Initiation events in in vitro packaging of bacteriophage φ29 DNA-p3. J. Mol. Biol. 197:219–228.
66. Guo, P., C. Peterson, and D. Anderson. 1987. Prohead and DNA-gp3-dependent ATPase activity of the DNA packaging protein gp16 of bacteriophage φ29. J. Mol. Biol. 197:229–236.
67. Guo, P., C. Zhang, C. Chen, K. Garver, and M. Trottier. 1998. Inter-RNA interaction of phage φ29 pRNA to form a hexameric complex for viral DNA transportation. Mol. Cell 2:149–155.
68. Gutiérrez, C., R. Freire, M. Salas, and J. M. Hermoso. 1994. Assembly of phage φ29 genome with viral protein p6 into a compact complex. EMBO J. 13:269–276.
69. Gutiérrez, C., G. Martín, J. M. Sogo, and M. Salas. 1991. Mechanism of stimulation of DNA replication by bacteriophage φ29 SSB protein p5. J. Biol. Chem. 266:2104–2111.
70. Gutiérrez, J., J. A. García, L. Blanco, and M. Salas. 1986. Cloning and template activity of the origins of replication of phage φ29 DNA. Gene 43:1–11.
71. Gutiérrez, J., J. Vinós, I. Prieto, E. Méndez, J. M. Hermoso, and M. Salas. 1986. Signals in the φ29 DNA-terminal protein template for the initiation of phage φ29 DNA replication. Virology 155:474–483.
72. Hagen, E. W., B. E. Reilly, M. E. Tosi, and D. L. Anderson. 1976. Analysis of gene function of bacteriophage φ29 of *Bacillus subtilis*: identification of cistrons essential for viral assembly. J. Virol. 19:501–517.
73. Harding, N. E., and J. Ito. 1980. DNA replication of bacteriophage φ29: characterization of the intermediates and location of the termini of replication. Virology 104:323–338.
74. Hendrix, R. W. 1998. Bacteriophage DNA packaging: RNA gears in a DNA transport machine. Cell 94:147–150.
75. Hermoso, J. M., E. Méndez, F. Soriano, and M. Salas. 1985. Location of the serine residue involved in the linkage between the terminal protein and the DNA of phage φ29. Nucleic Acids Res. 13:7715–7728.
76. Hoeprich, S., and P. Guo. 2002. Computer modeling of 3D structure of pRNA monomer, dimer and hexamer of φ29 DNA packaging motor. J. Biol. Chem. 277:20794–20803.
77. Horcajadas, J. A., W. J. J. Meijer, F. Rojo, and M. Salas. 2001. Analysis of early promoters of the *Bacillus* bacteriophage GA-1. J. Bacteriol. 183:6965–6970.
78. Horcajadas, J. A., M. Monsalve, F. Rojo, and M. Salas. 1999. The switch from early to late transcription in phage GA-1: characterization of the regulatory protein $p4_G$. J. Mol. Biol. 290:917–928.
79. Ibarra, B., J. L. Carrascosa, O. Llorca, M. Valle, and J. M. Valpuesta. 2000. Topology of the components of the DNA packaging machinery in the phage φ29 prohead. J. Mol. Biol. 298:807–815.
80. Illana, B., L. Blanco, and M. Salas. 1996. Functional characterization of the genes coding for the terminal protein and DNA polymerase from bacteriophage GA-1. Evidence for a sliding-back mechanism during protein-primed GA-1 DNA replication. J. Mol. Biol. 264:453–464.
81. Illana, B., J. M. Lázaro, C. Gutiérrez, W. J. J. Meijer, L. Blanco, and M. Salas. 1999. Phage φ29 terminal protein residues Asn80 and Tyr82 are recognition elements of the replication origins. J. Biol. Chem. 274:15073–15079.
82. Illana, B., A. Zaballos, L. Blanco, and M. Salas. 1998. The "RGD" sequence in phage φ29 terminal protein is required for interaction with φ29 DNA polymerase. Virology 248:12–19.

83. Inciarte, M. R. M. Salas, and J. M. Sogo. 1980. The structure of replicating DNA molecules of *Bacillus subtilis* phage ϕ29. J. Virol. 34:187–190.
84. King, A., W. R. Teertstra, and P. C. van der Vliet. 1997. Dissociation of the protein primer and DNA polymerase after initiation of adenovirus DNA replication. J. Biol. Chem. 272:24617–24623.
85. King, A.J, and P. C. van der Vliet. 1994. A precursor terminal protein-trinucleotide intermediate during initiation of adenovirus DNA replication: regeneration of molecular ends in vitro by a jumping back mechanism. EMBO J. 13:5786–5792.
86. Kornberg, A., and T. A. Baker. 1992. DNA Replication. W. H. Freeman, New York.
87. Martín, A. C., L. Blanco, P. García, M. Salas, and J. Méndez. 1996. Protein-primed initiation of pneumococcal phage Cp-1 DNA replication occurs at the third 3′ nucleotide of the linear template: a stepwise sliding-back mechanism. J. Mol. Biol. 260:369–377.
88. Martín, G., J. M. Lázaro, E. Méndez, and M. Salas. 1989. Characterization of the phage ϕ29 protein p5 as a single-stranded DNA binding protein. Function in ϕ29 DNA–protein p3 replication. Nucleic Acids Res. 17:3663–3672.
89. Matsumoto, K., T. Saito, and H. Hirokawa. 1983. In vitro initiation of bacteriophage ϕ29 and M2 DNA replication: genes required for formation of a complex between the terminal protein and 5′dAMP. Mol. Gen. Genet. 191:26–30.
90. Meijer, W. J. J., J. A. Horcajadas, and M. Salas. 2001. The ϕ29-family of phages. Microb. Mol. Biol. Rev. 65:261–287.
91. Meijer, W. J. J., P. J. Lewis, J. Errington, and M. Salas. 2000. Dynamic relocalization of phage ϕ29 DNA during replication and the role of the viral protein p16.7. EMBO J. 19:4182–4190.
92. Meijer, W. J. J., A. Serna-Rico, and M. Salas. 2001. Characterization of the bacteriophage ϕ29-encoded protein p16.7: a membrane protein involved in phage DNA replication. Mol. Microbiol. 39:731–746.
93. Mellado, R. P., F. Moreno, E. Viñuela, M. Salas, B. E. Reilly, and D. L. Anderson. 1976. Genetic analysis of bacteriophage ϕ29 of *Bacillus subtilis*: integration and mapping of reference mutants of two collections. J. Virol. 19:495–500.
94. Mellado, R. P., M. A. Peñalva, M. R. Inciarte, and M. Salas. 1980. The protein covalently linked to the 5′ termini of the DNA of *Bacillus subtilis* phage ϕ29 is involved in the initiation of DNA replication. Virology 104:84–96.
95. Mencía, M., M. Monsalve, F. Rojo, and M. Salas. 1996. Transcription activation by phage ϕ29 protein p4 is mediated by interaction with the α subunit of *Bacillus subtilis* RNA polymerase. Proc. Natl. Acad. Sci. USA 93:6616–6620.
96. Mencía, M., M. Monsalve, M. Salas, and F. Rojo. 1996. Transcriptional activator of phage ϕ29 late promoter: mapping of residues involved in interaction with RNA polymerase and in DNA bending. Mol. Microbiol. 20:273–282.
97. Méndez, J., L. Blanco, and M. Salas. 1997. Protein-primed DNA replication: a transition between two modes of priming by a unique DNA polymerase. EMBO J. 16:2519–2527.
98. Méndez, J., L. Blanco, J. A. Esteban, A. Bernad, and M. Salas. 1992. Initiation of ϕ29 DNA replication occurs at the second 3′ nucleotide of the linear template: a sliding-back mechanism for protein-primed DNA replication. Proc. Natl. Acad Sci. USA 89:9579–9583.
99. Méndez, J., L. Blanco, J. M. Lázaro, and M. Salas. 1994. Primer-terminus stabilization at the ϕ29 DNA polymerase active site: mutational analysis of conserved motif Tx_2GR. J. Biol. Chem. 269:30030–30038.
100. Monsalve, M., B. Calles, M. Mencía, F. Rojo, and M. Salas. 1998. Binding of phage ϕ29 protein p4 to the early A2c promoter: recruitment of a repressor by the RNA polymerase. J. Mol. Biol. 283:559–569.
101. Monsalve, M., B. Calles, M. Mencía, M. Salas, and F. Rojo. 1997. Transcription activation or repression by phage ϕ29 protein p4 depends on the strength of the RNA polymerase-promoter interactions. Mol. Cell 1:1–9.
102. Monsalve, M., M. Mencía, F. Rojo, and M. Salas. 1995. Transcription regulation in *Bacillus subtilis* phage ϕ29: expression of the viral promoters throughout the infection cycle. Virology 207:23–31.
103. Monsalve, M., M. Mencía, F. Rojo, and M. Salas. 1996. Activation and repression of transcription at two different phage ϕ29 promoters are mediated by interaction of the same residues of regulatory protein p4 with the RNA polymerase. EMBO J. 15:101–109.
104. Monsalve, M., M. Mencía, M. Salas, and F. Rojo. 1996. Protein p4 represses phage ϕ29 A2c promoter by interacting with the α subunit of *Bacillus subtilis* RNA polymerase. Proc. Natl. Acad. Sci. USA 93:8913–8919.
105. Moreno, F., A. Camacho, E. Viñuela, and M. Salas. 1974. Suppressor-sensitive mutants of phage ϕ29. Virology 62:1–16.
106. Müller, D. J., A. Engel, J. L. Carrascosa, and M. Vélez. 1997. The bacteriophage ϕ29 head–tail connector imaged at high resolution with the atomic force microscope in buffer solution. EMBO J. 16:2547–2553.
107. Nuez, B., and M. Salas. 1993. Bacteriophage Nf DNA region controlling late transcription: structural and functional homology with bacteriophage ϕ29. Nucleic Acids Res. 21:2861–2865.
108. Nuez, B., F. Rojo, and M. Salas. 1992. Phage ϕ29 regulatory protein p4 stabilizes the binding of the RNA polymerase to the late promoter in a process involving direct protein-protein contacts. Proc. Natl. Acad Sci. USA 89:11401–11405.
109. Paces, V., C. Vlcek, P. Urbanek, and Z. Hostomsky. 1986. Nucleotide sequence of the right early region of *Bacillus subtilis* phage PZA completes the 19366-bp sequence of PZA genome. Comparison with the homologous sequence of phage ϕ29. Gene 44:115–120.
110. Pastrana, R., J. M. Lázaro, L. Blanco, J. A. García, E. Méndez, and M. Salas. 1985. Overproduction and purification of protein p6 of *Bacillus subtilis* phage ϕ29: role in the initiation of DNA replication. Nucleic Acids Res. 13:3083–3100.
111. Pecenkova, T., and V. Paces. 1999. Molecular phylogeny of ϕ29-like phages and their evolutionary relatedness to other protein-primed replicating phages and other phages hosted by Gram-positive bacteria. J. Mol. Biol. 48:197–208.

112. Pecenkova, T., V. Benes, J. Paces, C. Vlcek, and V. Paces. 1997. Bacteriophage B103: complete DNA sequence of its genome and relatioship to other *Bacillus* phages. Gene 199:157–163.
113. Peñalva, M. A., and M. Salas. 1982. Initiation of phage ϕ29 DNA replication in vitro: formation of a covalent complex between the terminal protein, p3, and 5′-dAMP. Proc. Natl. Acad. Sci. USA 79:5522–5526.
114. Prieto, I., E. Méndez, and M. Salas. 1989. Characterization, overproduction and purification of the product of gene 1 of *Bacillus subtilis* phage ϕ29. Gene 77:195–204.
115. Prieto, I., M. Serrano, J. M. Lázaro, M. Salas, and J. M. Hermoso. 1988. Interaction of the bacteriophage ϕ29 protein p6 with double-stranded DNA. Proc. Natl. Acad. Sci. USA 85:314–318.
116. Rojo, F., and M. Salas. 1991. A DNA curvature can substitute phage ϕ29 regulatory protein p4 when acting as a transcriptional repressor. EMBO J. 10:3429–3438.
117. Rojo, F., M. Mencía, M. Monsalve, and M. Salas. 1998. Transcription activation and repression by interaction of a regulator with the α subunit of RNA polymerase: the model of phage ϕ29 protein p4. Prog. Nucleic Acids Res. Mol. Biol. 60:29–46.
118. Saedi, M. S., K. J. Garvey, and J. Ito. 1987. Cloning and purification of a unique lysozyme produced by *Bacillus* phage ϕ29. Proc. Natl. Acad. Sci. USA 84:955–958.
119. Salas, M. 1988. Phages with protein attached to the DNA ends. pp. 169–191. *In* R. Calendar (ed.) The Bacteriophages, vol. I. Plenum Press, New York.
120. Salas, M. 1991. Protein-priming of DNA replication. Annu. Rev. Biochem. 60:39–71.
121. Salas, M. 1999. Bacillus phage ϕ29, pp. 119–130. *In* R. G. Webster and A. Granoff (eds.) Encyclopedia of Virology, 2nd ed, vol. 1. Saunders, Philadelphia.
122. Salas, M., J. Méndez, J. A. Esteban, M. Serrano, C. Gutierrez, J. M. Hermoso, A. Bravo, M. S. Soengas, J. M. Lázaro, M. A. Blasco, R. Freire, A. Bernad, J. M. Sogo, and L. Blanco. 1993. Terminal protein priming of DNA replication: bacteriophage ϕ29 as a model system. Pp. 3–19. *In* W. Doerfler and P. Böhm (eds.) Virus Strategies, Molecular Biology and Pathogenesis, Verlag Chemie, Weinheim.
123. Salas, M., J. T. Miller, J. Leis, and M. L. DePamphilis. 1996. Mechanisms for priming DNA synthesis. pp. 131–176. *In* M. DePamphilis (ed.) DNA Replication in Eukaryotic Cells. Cold Spring Harbor Laboratoyr Press, Cold Spring Harbor, N.Y.
124. Saturno, J., L. Blanco, M. Salas, and J. A. Esteban. 1995. A novel kinetic analysis to calculate nucleotide affinity of proofreading DNA polymerases. Application to ϕ29 DNA polymerase fidelity mutants. J. Biol. Chem. 270:1–9.
125. Saturno, J., J. M. Lázaro, L. Blanco, and M. Salas. 1998. Role of the first aspartate of the "YxDTDS" motif of ϕ29 DNA polymerase as a metal ligand during both TP-primed and DNA-primed DNA synthesis. J. Mol. Biol. 283:633–642.
126. Saturno, J., J. M. Lázaro, F. J. Esteban, L. Blanco, and M. Salas. 1997. ϕ29 DNA polymerase residue Lys 383, invariant at motif B of DNA dependent polymerases, is involved in dNTP binding. J. Mol. Biol. 269:313–325.
127. Serna-Rico, A., B. Illana, M. Salas, and W. J. J. Meijer. 2000. The putative coiled coil domain of the ϕ29 terminal protein is a major determinant involved in recognition of the origin of replication. J. Biol. Chem. 275:40529–40538.
128. Serna-Rico, A., M. Salas, and W. J. J. Meijer. 2002. The *Bacillus subtilis* phage ϕ29 protein p16.7, involved in ϕ29 DNA replication, is a membrane-localized single-stranded DNA-binding protein. J. Biol. Chem. 277:6733–6742.
129. Serrano, M., C. Gutiérrez, R. Freire, A. Bravo, M. Salas, and J. M. Hermoso. 1994. Phage ϕ29 protein p6: a viral histone-like protein. Biochimie 76:981–991.
130. Serrano, M., C. Gutiérrez, M. Salas, and J. M. Hermoso. 1993. Superhelical path of the DNA in the nucleoprotein complex that activates the initiation of phage ϕ29 DNA replication. J. Mol. Biol. 230:248–259.
131. Serrano, M., J. Gutiérrez, I. Prieto, J. M. Hermoso, and M. Salas. 1989. Signals at the bacteriophage ϕ29 DNA replication origins required for protein p6 binding and activity. EMBO J. 8:1879–1885.
132. Serrano, M., M. Salas, and J. M. Hermoso. 1990. A novel nucleoprotein complex at a replication origin. Science 248:1012–1016.
133. Simpson, A. A., Y. Tao, P. G. Leiman, M. O. Badasso, Y. He, P. J. Jardine, N. H. Olson, M. C. Morais, S. Grimes, D. L. Anderson, T. S. Baker, and M. G. Rossmann. 2000. Structure of the bacteriophage ϕ29 packaging motor. Nature 408:745–750.
134. Smith, D. E., S. J. Tans, S. B. Smith, S. Grimes, D. L. Anderson, and C. Bustamante. 2001. The bacteriophage ϕ29 portal motor can package DNA against a large internal force. Nature 413:748–752.
135. Soengas, M. S., J. A. Esteban, J. M. Lázaro, A. Bernad, M. A. Blasco, M. Salas, and L. Blanco. 1992. Site-directed mutagenesis at the Exo III motif of ϕ29 DNA polymerase. Overlapping structural domains for the 3′→5′ exonuclease and strand-displacement activities. EMBO J. 11:4227–4237.
136. Soengas, M. S., J. A. Esteban, M. Salas, and C. Gutiérrez. 1994. Complex formation between phage ϕ29 single-stranded DNA binding protein and DNA. J. Mol. Biol. 239:213–226.
137. Soengas, M. S., C. Gutiérrez, and M. Salas. 1995. Helix-destabilizing activity of ϕ29 single-stranded DNA binding protein: effect on the elongation rate during strand displacement DNA replication. J. Mol. Biol. 253:517–529.
138. Steiner, M., W. Lubitz, and U. Bläsi. 1993. The missing link in phage lysis of Gram-positive bacteria: gene 14 of *Bacillus subtilis* phage ϕ29 encodes the functional homolog of lambda S protein. J. Bacteriol. 175:1038–1042.
139. Talavera, A., F. Jiménez, M. Salas, and E. Viñuela. 1971. Temperature sensitive mutants of bacteriophage ϕ29. Virology 46:586–595.
140. Talavera, A., M. Salas, and E. Viñuela. 1972. Temperature-sensitive mutants affected in DNA synthesis in phage ϕ29 of *Bacillus subtilis*. Eur. J. Biochem. 31:367–371.

141. Tao, Y., N. H. Olson, W. Xu, D. L. Anderson, M. G. Rossmann, and T. S. Baker. 1998. Assembly of a tailed bacterial virus and its genome release studied in three dimensions. Cell 95:431–437.

142. Tedin, K., A. Resch, M. Steiner, and U. Bläsi. 1995. Dual translation start motif evolutionarily conserved in the holin gene of *Bacillus subtilis* phage ϕ29. Virology 206:479–484.

143. Trottier, M., C. L. Zhang, and P. Guo. 1996. Complete inhibition of virion assembly in vivo with mutant pRNA essential for phage ϕ29 DNA packaging. J. Virol. 70:55–61.

144. Truniger, V., L. Blanco, and M. Salas. 1999. Role of the "YxGG/A" motif of ϕ29 DNA polymerase in protein-primed replication. J. Mol. Biol. 286:57–69.

145. Truniger, V., J. M., Lázaro, M. Salas, and L. Blanco. 1996. A DNA binding motif coordinating synthesis and degradation in proofreading DNA polymerases. EMBO J. 15:3430–3441.

146. Truniger, V., J. M. Lázaro, M. Salas, and L. Blanco. 1998. ϕ29 DNA polymerase requires the N-terminal domain to bind terminal protein and DNA primer substrates. J. Mol. Biol. 278:741–755.

147. Valpuesta, J. M., J. M. Carazo, and J. L. Carrascosa. 1999. The three-dimensional structure of a DNA translocating machine at 10 Å resolution. Structure 7:289–296.

148. Valle, M., L. Kremer, C. Martínez-A., F. Roncal, J. M. Valpuesta, J. P. Albar, and J. L. Carrascosa. 1999. Domain architecture of the bacteriophage ϕ29 connector protein. J. Mol. Biol. 288:899–909.

149. Vlcek, C., and V. Paces. 1986. Nucleotide sequence of the late region of *Bacillus* phage ϕ29 completes the 19285-bp sequence of ϕ29. Comparison with the homologous sequence of phage PZA. Gene 46:215–225.

150. Watabe, K., M. Leusch, and J. Ito. 1984. Replication of bacteriophage ϕ29 DNA in vitro: The roles of terminal protein and DNA polymerase. Proc. Natl. Acad. Sci. USA 81:5374–5378.

151. Watabe, K., M. Leusch, and J. Ito. 1984. A 3′ to 5′ exonuclease activity is associated with phage ϕ29 DNA polymerase. Biochem. Biophys. Res. Commun. 123: 1019–1026.

152. Whiteley, H. R., W. D. Ramey, G. B. Spiegelman, and R. D. Holder. 1986. Modulation of in vivo and in vitro transcription of bacteriophage ϕ29 early genes. Virology 155:392–401.

153. Yoshikawa, H., K. J. Garvey, and J. Ito. 1985. Nucleotide sequence analysis of DNA replication origins of the small bacteriophage: evolutionary relationship. Gene 37:125–130.

154. Zaballos, A. and M. Salas. 1989. Functional domains in the bacteriophage ϕ29 terminal protein for interaction with the ϕ29 DNA polymerase and with DNA. Nucleic Acids Res. 17:10353–10366.

155. Zhang, F., S. Lemieux, X. Wu, D. St.-Arnaud, T. C. McMurray, F. Major, and D. Anderson. 1998. Function of hexameric RNA in packaging of bacteriophage ϕ29 DNA in vitro. Mol. Cell 2:141–147.

23

Bacteriophage SPP1

JUAN C. ALONSO
PAULO TAVARES
RUDI LURZ
THOMAS A. TRAUTNER

The lytic *Bacillus subtilis* bacteriophage SPP1 was isolated in Pavia (Subtilis Phage Pavia) and first described by Riva et al. (99). Interest in this phage derived at first from the observation that the complementary DNA strands of the phage were separable into a heavy strand ($\rho = 1.725$ g/cm^3) and a light strand ($\rho = 1.713$ g/cm^3) following denaturation and isopycnic centrifugation (99, 109). This observation and the finding that the phage was highly active in transfection led to the use of SPP1 in studies of mismatch repair (109). In later experiments the molecular biology of the phage, including its structure, total DNA sequence, genetics, and life cycle were established. Furthermore, SPP1 transfection was a convenient tool to study the mechanism of DNA processing during uptake into competent cells (123). SPP1 is a generalized transducing phage with respect to bacterial host markers and it also transduces plasmid DNA. These studies made SPP1 one of the best characterized bacteriophages of Gram-positive bacteria. In this review we shall summarize the results obtained in our and other laboratories during the last 35 years. Special attention is given to comparisons of the results obtained with SPP1 with properties of other organisms, providing a contribution to the understanding of the evolution of bacterial viruses. It should be realized that all findings reported here were obtained under laboratory conditions, which would not necessarily reflect conditions which the phage would encounter in its natural habitat.

General Properties

Properties of the Virion

The SPP1 virion has an isometric icosahedral capsid of approximately 60 nm diameter attached to a long noncontractile tail composed of a helical tube (177 nm long) and a fiber responsible for host cell attachment (99; figure 23-1). SPP1 belongs to the morphotype B1 of the *Siphoviridae* family (1, 18; chapter 2). SPP1 is the prototype of the proposed genus “SPP1-like viruses” that includes phages ρ15, SF6, and 41c.

The standard host for SPP1 is *B. subtilis* strain 168 and its derivatives. On this strain and using tryptone-yeast (TY) plates (15, 100) SPP1 wild-type produces plaques with a clear center of about 2 mm and a halo with a width of about 1 mm during overnight incubation. Numerous mutants affecting the plaque type, temperature sensitivity, and growth on suppressor-sensitive strains have been identified. Phage stocks with titers of 10^{10} to 10^{11} are readily obtained by infection of host bacteria growing in liquid TY cultures or in synthetic medium. Phage stocks can be concentrated further by precipitation with polyethylene glycol. Virions are very stable provided they are maintained in the presence of 10^{-2} M Mg^{2+}, Ca^{2+}, or Mn^{2+} ions. Chelating agents such as citrate, pyrophosphate, EGTA, or EDTA rapidly destroy infectivity. Adsorption-resistant mutants of *B. subtilis* 168 have been isolated and the associated gene was mapped in the host genome. The nature of the bacterial receptor for SPP1 adsorption has not been identified. Phage SPP1 cannot stably establish itself in infected cells as a prophage. Phage stocks, however, contain a fraction of transducing particles with only bacterial DNA, which will deliver such DNA to the recipient host bacterium without lysing the cell. SPP1 is sensitive to restriction by some of the restriction/modification systems identified in *B. subtilis* (122). Infections of SPP1 and other *B. subtilis* phages do not produce mixed bursts. However, SPP1 forms viable hybrid virions in mixed infections with its highly related phages ρ15, SF6, and 41c (101, 104). Multiple infection with *several* SPP1 phages per cell affects neither the burst size nor the latent period. Such infections permit the performance of genetic crosses. Crosses using a large

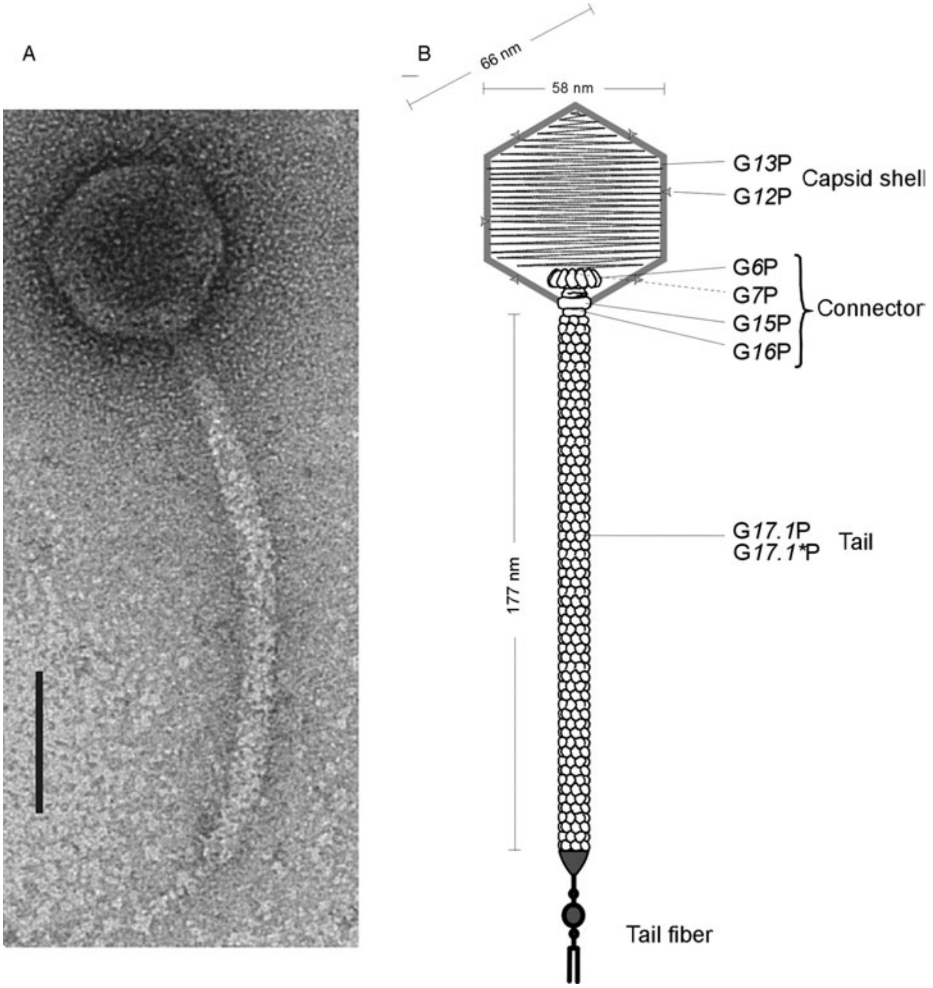


Figure 23-1 Structure of bacteriophage SPP1. A: Visualization of an SPP1 virion by electron microscopy after negative staining with uranyl acetate. The bar represents 50 nm. B: a scheme of the mature phage that compiles our current knowledge of the particle structural organization. The icosahedral capsid is formed by the main capsid protein G*13*P and a decoration protein G*12*P (12). The connector between capsid and phage tail is composed of the portal protein G6P and the head completion proteins G*15*P and G*16*P (12, 76). The location of the minor capsid protein G7P in the mature phage head has not been identified but the protein is known to bind G6P at early stages of SPP1 morphogenesis (45, 115). The two main structural proteins of the tail are G*17.1*P and G*17.1**P, which have an identical N-terminus but G*17.1**P has a higher molecular weight (43). The structure of the tail fiber is drawn based on visual inspection of electron micrographs. Its individual protein components have not been defined.

diversity of mutant phages have led to the establishment of a linear genetic map of SPP1 with distances between markers given in probabilities of recombination between them (110).

DNA

SPP1 DNA can be readily obtained and purified by conventional methods. SPP1 wild type DNA has a contour length of about 15 μm. The molecules are partially circularly permuted and have a terminal redundancy of 4% (89, 119). In most of our previous communications, as well as in figures 23-2 and 23-3, SPP1 DNA is represented as the ensemble of ordered *Eco*RI restriction fragments (13, 97, 104). DNA of *Eco*RI fragment 13 and of the major part of *Eco*RI fragment 1 are dispensable following the identification of deletions in these regions. A derivative of SPP1 with an inversion of some 8.4 kb, extending from within *Eco*RI restriction fragment 1 to restriction fragment 5, was generated during a cloning experiment (41). SPP1 DNA does not contain any unusual bases. Properties of the DNA determined by physicochemical or chemical methods were refined and confirmed following the establishment of the total base sequence of SPP1 DNA (3). According to this sequence the SPP1 genome has a size of 44,007 bp and a base composition of 43.7% dG + dC. The compiled sequence of SPP1 DNA is available under EMBL accession number X97918.

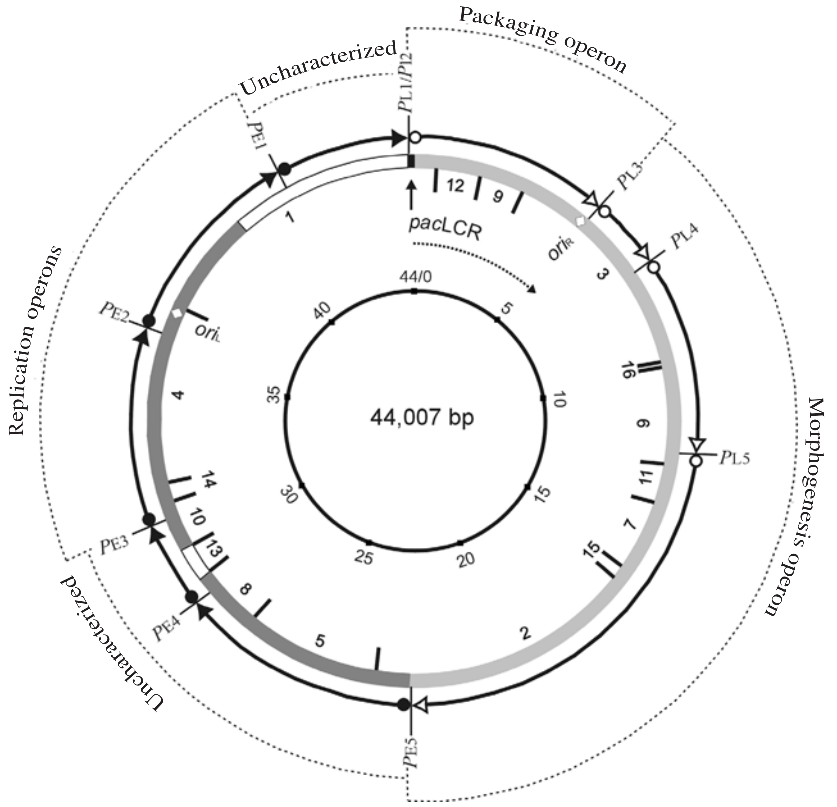


Figure 23-2 Physical and genetic map of the genome of bacteriophage SPP1. The genome is presented in a circular form as found in the cell at early times after infection. The inner circle represents scale from 1 to 44,007 base pairs (3). Digestion of the DNA with *Eco*RI generates 16 fragments as indicated by lines and numbers in the thick bar circle. The packaging initiation sequence *pac*LCR (see section "Morphogenesis") is shown as a black box indicated by a vertical arrow. The direction of packaging is shown by a dotted arrow. The white rhombi within the *Eco*RI-4 and *Eco*RI-3 fragments denote the *ori*L and *ori*R sequences, respectively (see section "DNA Replication"). Segments of the genome dispensable for SPP1 multiplication are shown in white frames. The region coding for genes expressed early during infection is shown in dark grey and the region coding for late genes in light grey. The five early promoters (*P*E1–*P*E5) are indicated by filled circles in the outer periphery of the map. Arrows show the direction of transcription by σ^A-RNAP and arrowheads the position of rho-independent transcription terminators. The late σ^A-RNAP promoters (*P*L1–*P*L5; open circles) and their transcripts are represented using the same symbols. The operons coding for genes involved in DNA replication, DNA packaging, and virion morphogenesis, as well as the regions of the genome not characterized yet, are indicated on the outside. See thebacteriophages.org/frames_0230.htm for color version of this figure.

Proteins

The proteins encoded by SPP1 were first analyzed by SDS polyacrylamide gel electrophoresis following disruption of radioactively labeled phage particles or after lysis of SPP1 infected *B. subtilis* cells (51). The proteins identified could be subdivided into three groups on the basis of their synthesis during the latent period. Such analyses were complicated by the fact that SPP1 infection does not interfere with host protein synthesis. Also, neither the irradiation of cells prior to infection nor the addition of drugs that were anticipated to selectively inhibit host protein synthesis provided conditions for the study of phage-coded protein synthesis (49). To eliminate the background of host protein synthesis, further characterization of SPP1 proteins included the expression of SPP1 genes in infected minicells of *B. subtilis* (83) or in *Escherichia coli* infected with λ/SPP1 hybrid phages which, together, contained all clonable *Eco*RI restriction fragments of SPP1 DNA (4, 5). The set of proteins potentially encoded by SPP1 is shown in figure 23-3 (3). These include reading frames (ORFs) coding for functionally identified proteins, for proteins which are dispensable, and for unidentified proteins that remain to be characterized.

Life Cycle

Adsorption and Infection

SPP1 infection is initiated by adsorption of the phage to *B. subtilis* 168. As is the case for numerous other

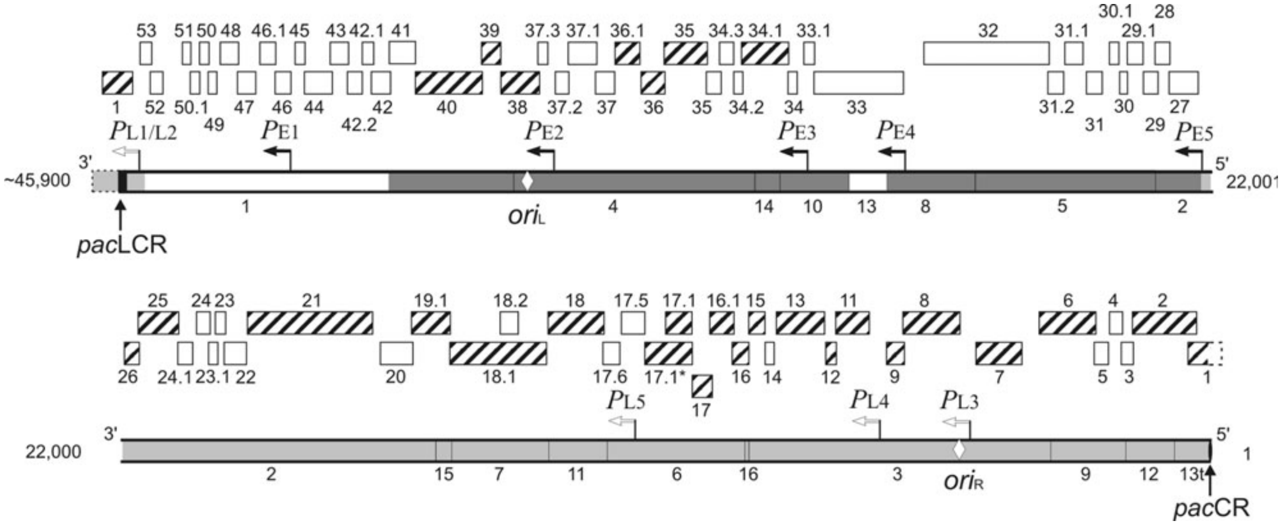


Figure 23-3 The bacteriophage SPP1 mature chromosome. The two main bars denote the mature phage sequence (1 to ~45.9 kb). The upper border of the bar represents the H (heavy) strand and the lower line the L (light) strand. Also indicated are the 16 *Eco*RI DNA fragments (numbered in order of decreasing size), the dispensable, early and late gene regions, the early (P_E) and late (P_L), and the origins of replication ori_L and ori_R (symbols are as in figure 23-2). Fragment *Eco*RI-13t has one end generated by *Eco*RI digestion and the other by cleavage at the *pac* site by the terminase enzyme. The *pac* region is located within *Eco*RI DNA fragment 1 (uncleaved: *pac*LCR, filled black box indicated by a vertical arrow). The cleaved *pac* site (*pac*CR, filled black oval, indicated by a vertical arrow) defines the first nucleotide of the SPP1 sequence. To present the full *Eco*RI DNA fragment 1, the first 718 base pairs (defining the *Eco*RI-13t fragment) are shown again in the upper main bar after position 44007 (broken line). Therefore, gene *1* is also presented twice. Due to imprecision of the headful cut mechanism, the end of the mature SPP1 chromosome (~45.9 kb) is not precisely defined. The numbered hatched boxes denote phage-encoded products that have been genetically or biochemically characterized, whereas numbered but otherwise empty boxes indicate those of unknown function (3). See thebacteriophages.org/frames_0230.htm for color version of this figure.

bacteriophages, two different steps can be distinguished in SPP1 adsorption. The initial interaction of SPP1 with the host cell surface is reversible and infective virions can be detached from the host. Lipoteichoic acids, which are components of the Gram-positive cell wall, play a role in this step (101, M. A. Santos, personal communication). The second step is the irreversible binding of SPP1 to the host bacterium followed by DNA ejection. Host mutations preventing irreversible adsorption were mapped by cotransduction experiments to the locus *pha-2*, which is linked to *ald-1* in the *B. subtilis* chromosome (position 3277 kb; 71) (102). The SPP1 receptor was recently shown to be the transmembrane protein YueB (see note added in proof on p. 345). Mutations in *pha-2* confer resistance uniquely to infection by SPP1 and to its close relatives ρ15, SF6, and 41c. By contrast, resistance to other *B. subtilis* phages is associated with defects in the glucosylation of teichoic acids (*gta* mutations; 135). SPP1 and its relatives are some of the few phages known to infect *gta* strains (67). The SPP1-like phages thus appear to have developed a group-specific mechanism for delivery of DNA to the *B. subtilis* cytoplasm.

At low multiplicity of infection, $\geq 90\%$ of the phage DNA is transferred to the host cell interior within 10 minutes of phage–bacteria incubation (136). The phage chromosome becomes accessible to DNase I attack during its transfer to the host cell, demonstrating that it is exposed to the exterior medium at a stage of the transfer process (103, 136).

Delivery of the SPP1 chromosome to the *B. subtilis* cytoplasm leads to the successful recruitment of the host cell machineries necessary for high-level expression of the viral genes, to replicate the SPP1 DNA, and to support assembly of virions at high efficiency. This occurs in absence of host gene-expression shutoff (49).

Transcription

Transcription of SPP1 DNA is asymmetric and therefore unidirectional (figure 23-2; 50, 88, 98). Of the two separable strands of SPP1 DNA only the heavy strand, with 58.3% purine, is transcribed. This is, however, not an inherent requirement for the viability of SPP1: In the SPP1 mutant described before in which the entire region between *Eco*RI restriction fragments 1 and 5 has been inverted, transcription of early genes must occur from the DNA light strand (41).

In vivo and in vitro transcription studies located five early (*P*E1 to *P*E5) and five late (*P*L1 to *P*L5) promoters in the SPP1 genome (32, 33, 86, 95, 116, 117; figures 23.2, 23.3). The messenger RNAs are polycistronic. σ^A-RNA polymerase

(σ^A-RNAP) is utilized for the transcription of early SPP1 promoters (84). Similarly to strongly transcribed housekeeping genes, the five early promoters (*P*E1–*P*E5) have upstream of their consensus sequences an alternating poly-purine poly-pyrimidine (Pur.Pyr) region and a very high dA + dT content (AT-rich) (3, 63).

*P*E1 maps upstream of *orf45.1* (defining a nonessential operon), *P*E2 upstream of *orf37.3* (replication initiation operon), *P*E3 upstream of *orf34* (replication accessory proteins operon), and *P*E4 and *P*E5 upstream of *orf33* and of *orf27*, respectively (uncharacterized operons) (33, 95, 117; figures 23.2, 23.3). Transcription originating from *P*E1 and *P*E4 yields products that are dispensable for phage growth under laboratory conditions. Furthermore, some of the products read from the *P*E2 and *P*E5 promoters are also dispensable. Transcription from the *P*E1 promoter yields a messenger RNA that allows expression of *orf46* to *orf53*, but SPP1 mutants with a deletion in this segment do not appear to be in any way defective, provided transcription read-through into the late operon does not occur (see 33). Deletion of the central part of *orf33*, that is, the only product transcribed from *P*E4, does not affect the phage physiology (41; figures 23.2, 23.3).

The early promoters show a hierarchy of signal strength: when the promoter strength of *P*E2 is considered as 1, *P*E3 is 1.5–1.9 times stronger, whereas the strength of *P*E1, *P*E4, and *P*E5 is 0.8-, 0.6-, and 0.5-fold respectively that of *P*E2 (33, 116, 117). The presence of additional weak promoters located within the genes cannot be excluded.

The late genes, whose expression requires preceding phage-DNA synthesis (49), are defined by the five late operons [gene *1–7* operon (read from *P*L1 and *P*L2), gene *8–9* operon (*P*L3), gene *11–17.3* operon (*P*L4) and gene *17.5–26* (*P*L5)] (31–33). SPP1 late transcription, in addition to a replicating DNA substrate, requires an uncharacterized phage-encoded product (31). Unlike the case of phage λ, in which late transcription is activated by allowing the host RNAP to proceed through termination sites, the SPP1 transcriptional activator promotes late transcription by a different but uncharacterized mechanism (see 3). The best-characterized SPP1 late promoters are *P*L1 and *P*L2 (31–33). The *P*L1 promoter, which accounts for about 95% of the transcripts of the gene *1–7* operon, lacks the −35 σ^A-RNAP consensus region, whereas the weak *P*L2 promoter, whose initiation site maps 35 nucleotides downstream of the *P*L1 start site, has all the features present in a vegetative *B. subtilis* promoter (32). Furthermore, transcription within the gene *1* operon, which is read from the *P*L1 and *P*L2 promoters, is shut off by the terminase enzyme at late times (18 minutes) after infection (31).

DNA Replication

Replication of the SPP1 DNA is mediated at least partly by the host DNA polymerase III. This follows from the observation that SPP1 DNA synthesis is sensitive to the drug HPUra (49) and that mutations are induced in SPP1 when the phage grows on host cells with a mutator polymerase III (87).

SPP1 packs its double-stranded (ds) DNA into an empty procapsid by a processive headful mechanism, using a linear head-to-tail concatemer as a substrate (30, 32, 59, 89). SPP1 circular molecules were detected in infected cells, but branched replication intermediates have not been observed (24, 56, 80, 81). Furthermore, a block in DNA replication does not reveal the accumulation of head-to-tail concatemers (24, 81). Initiation of ring-to-ring replication with the subsequent shift via recombination-dependent DNA replication (RDR) leads to the generation of concatemers that are the substrate for DNA packaging (see 107, 131).

Analysis of SPP1 conditional-lethal mutants for their capacity to synthesize phage DNA has led to the identification of five different complementation groups. Mutants in genes *38*, *39*, and *40* show a block in DNA replication, whereas mutants in genes *34.1* and *35* show a normal initiation but a replication arrest phenotype (24, 95, 132). Accumulation of ring-to-ring (theta replication) SPP1 replication intermediates has not been observed in SPP1-infected cells (56, 80, 81).

Initiation of SPP1 Theta Replication

SPP1 ring-to-ring replication begins at a "unique" origin and proceeds unidirectionally (56, 81). Previously it has been shown that G*38*P,* G*39*P and G*40*P are the only SPP1-encoded functions necessary and sufficient to drive theta replication from the *cis*-acting *ori*L region in an otherwise nonreplicative element in *B. subtilis* cells (85). An SPP1 sequence termed *ori*R has also been located 12 kb away from *ori*L in the SPP1 circular map (24, 56, 81, 85, 95; figures 23.2, 23.3). In vitro studies have revealed that multiple copies of G*38*P bound to its cognate site induce local unwinding of the adjacent AT-rich sequence present within *ori*L or *ori*R (85; figure 23-4).

Initiation of SPP1 DNA replication strictly requires G*38*P, G39P, and G*40*P as well as the host DNA primase (DnaG), DNA Pol III and topoisomerases of types I and II. In the absence of ATP, multiple copies of monomeric G*38*P, bound to its cognate sites on a supercoiled substrate, induce local unwinding of the adjacent AT-rich sequence, leading to open complex formation (6, 7, 85, 95). The single-stranded binding protein (SSB-G*36*P) binds to the ssDNA at the open complex (figure 23-4). G*39*P is a natural partially unfolded protein that is folded upon

---

* The designations G*X*P (gene *X* product) and gpX (gene product X) are used for the same SPP1 protein. We follow the first terminology in this review.

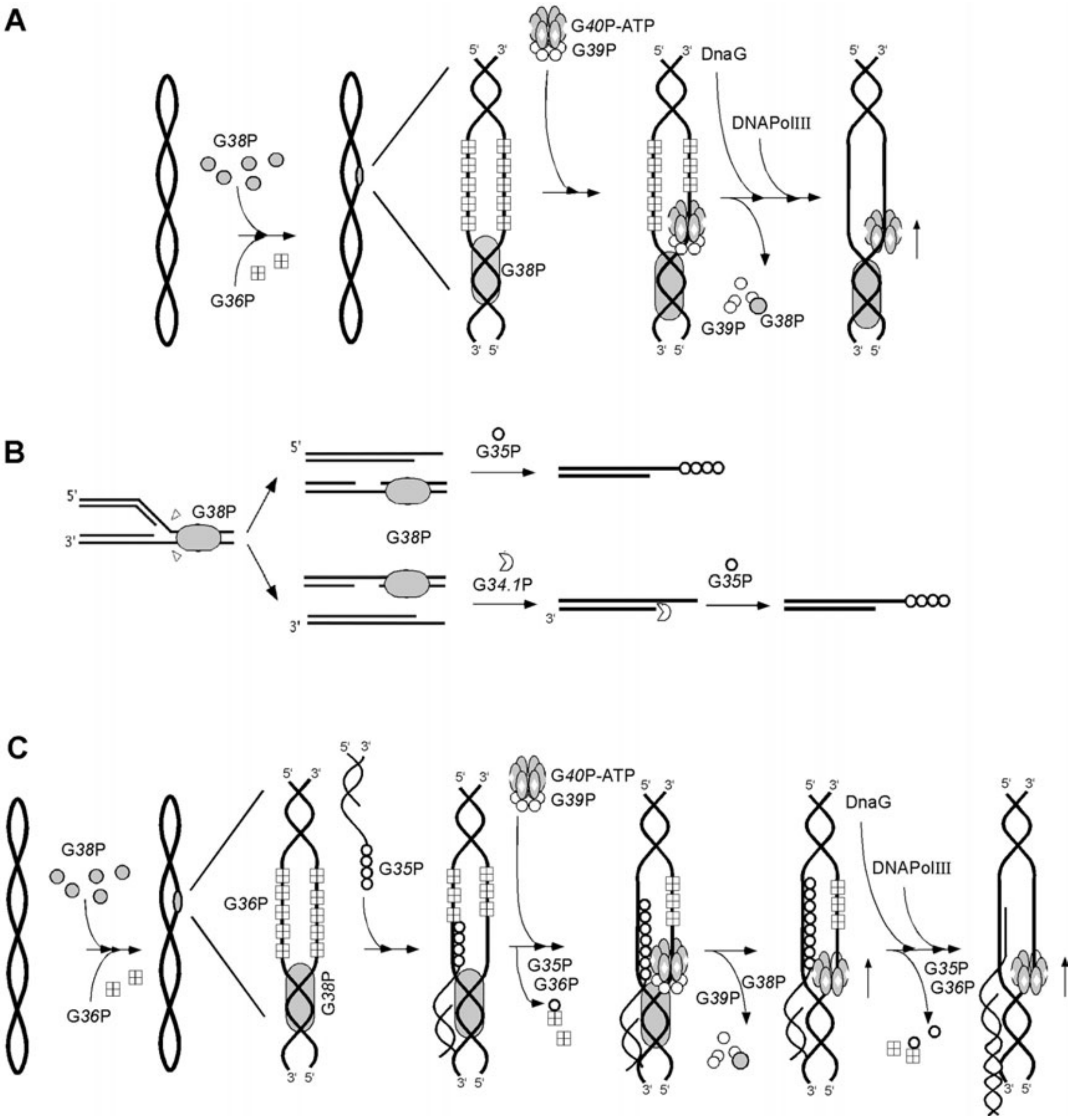


Figure 23-4 Model for phage SPP1 initiation of DNA replication. A: Model for SPP1 initiation of theta type DNA replication. First, G*38*P recognizes AB boxes of the SPP1 replication origin (*ori*L), and binds to them in an ATP-independent fashion. Binding of G*38*P leads to melting of the AT-rich region adjacent to the AB boxes. The helix destabilizer, G*36*P, binds to the signle-stranded DNA. The G*39*P–G*40*P–ATP complex is loaded in the unwound region by the interactions between G*39*P and G*38*P and the ATP-dependent single-strand DNA binding capacity of G*40*P. The interaction of G*39*P with G*38*P, which form a G*38*P–G*39*P heterodimer, dissociates G*39*P from the G*39*P–G*40*P–ATP complex and the helicase is free of the inhibitory effect of its loader with subsequent release of the G*39*P monomers and homodimers and G*38*P–G*39*P heterodimers from the nucleoprotein complex. Both DnaG and DNA pol III upon interacting with G*40*P are then loaded at the origin. The hydrolysis of ATP would then produce the unwinding of the DNA by G*40*P and its translocation. B: Roadblock as a model for the shift from theta to sigma replication. G*38*P bound to *ori*L or *ori*R blocks replication fork progression. A nick in the leading strand (bottom strand) will be processed by the putative 5′–3′ exonuclease, G*34.1*P, to generate a 3′ single-strand DNA tail on which G*35*P will polymerize. A nick in the lagging strand (top strand) will not require further processing, and G*35*P will polymerise on it. C: Model for SPP1 initiation of sigma type DNA replication. Binding of G*38*P, with the help of G*36*P, stabilizes the melted AT-rich region adjacent to the AB boxes at either *ori*L or *ori*R. A G*35*P-single-strand DNA filament pairs with the complementary strand of the unwound region to form a D-loop. The G*39*P–G*40*P–ATP complex is then loaded in the unwound region by the interactions between G*39*P and G*38*P and the ATP-dependent single-stranded DNA binding capacity of G*40*P. The interaction of G*39*P with G*38*P, leading to G*39*P monomers and homodimers and G*38*P–G*39*P heterodimers, dissociates G*39*P from the G*39*P–G*40*P–ATP complex. G*40*P helps the assembly of both DnaG and DNA polymerase III at the AT-rich region. The 3′-OH end of the paired strand could be used to prime the leading strand and DnaG could provide the primer for lagging-strand synthesis. See thebacteriophages.org/frames_0230.htm for color version of this figure.

interaction with G*40*P-ATP and which shuts off all activities associated with the latter (e.g., ATPase, helicase, ssDNA binder) (7, 9). G*40*P-ATP DNA helicase, which is a hexamer in solution, interacts with G*39*P (7, 10). Both proteins form a heterododecamer (G*39*P_6–G*40*P_6–ATP) in solution. G*39*P of the G*39*P_6–G*40*P_6–ATP complex interacts with G*38*P bound to *ori*L, loads G*40*P_6–ATP onto the ssDNA at the open complex, and forms a heterodimer with G*38*P (figure 23-4). Then, the G*38*P–G*39*P complex is released from the replication complex (7). G*40*P–ATP released from the G*39*P_6–G*40*P_6–ATP complex is "activated". G*40*P_6–ATP is stabilized on the ssDNA upon interaction with SSB–G*36*P bound to the complementary strand and interacts with the host-encoded monomeric DnaG (6) and with the τ subunit of DNA Pol III (79), and SPP1 theta type DNA replication initiates (see figure 23-4).

Initiation of SPP1 Sigma Replication

The molecular events that trigger the synthesis of head-to-tail concatemeric SPP1 DNA are not well understood. The picture of how RDR proceeds has changed over the last decade (35, 72, 78, 131). SPP1 replication fork re-start, which leads to the accumulation of SPP1 concatemers, is independent of host-encoded RecA, AddAB, and RecF (counterpart of *E. coli* RecA, RecBCD, and RecF, respectively) recombination proteins, of the replicative DNA helicase and its loader (DnaC and DnaI, respectively), and of components of the *B. subtilis* primosome (e.g., PriA, DnaB, DnaD) (2, 8, 80, 95, 132).

Genetic data suggest that the generation of linear concatemeric SPP1 DNA requires the phage-encoded G*35*P and G*34.1*P proteins. Sedimentation studies of DNA synthesized by SPP1*tsI20F* (impaired in G*34.1*P) or SPP1*tsI17* (impaired in gene *35*) infected cells, at the restrictive temperature, revealed that only a small percentage of the phage DNA can be recovered in a fast sedimenting form (concatemeric DNA). In both cases less-than-unit-length SPP1 genomes (30–35 kb in size) accumulated (24). Considering the unidirectional movement of the SPP1 replication fork we can assume that any event initiated at *ori*L that stops at *ori*R will generate 32 kb replication intermediates. G*35*P binds and filaments on ssDNA, catalyzes, in an ATP-independent manner, joint molecule formation between a 3′-ssDNA and a homologous AT-rich region of *ori*L on a supercoiled molecule, and specifically interacts with the G*40*P DNA helicase and G*36*P SSB proteins (8). G*34.1*P shares 18% identity with the ATP-independent 5′ → 3′ exonucleases Rac-encoded RecE product (also termed ExoVIII). After Skalka (107), Formosa and Alberts (54), Viret et al. (131), and Kuzminov (72) we hypothesize that after the initial phase of initiation of theta replication at *ori*L, the progression of the replication fork might be stalled when it encounters G*38*P bound at the inversely oriented *ori*R (roadblock) in the absence of overt DNA damage (85, 95; figure 23-4B), or at any region in the presence of DNA damage. The former claim is consistent with the observation that progression of a replicating fork in vivo is transiently stalled when the replisome approaches an inversely oriented silent origin on a plasmid (see 129). The stalled replication fork breaks and the broken fork is rescued by a process dependent on phage-encoded G*34.1*P and G*35*P functions since such a defect cannot be overcome by any of the host recombination and/or PriA-dependent replication functions (2, 8, 80, 95, 132). The G*34.1*P exonuclease may generate a duplex DNA with a 3′ terminus (Martínez-Jiménez, personal communication; figure 23-4B). G*35*P-mediated joint molecule formation could provide a 3′- end to anneal at *ori*R on a second supercoiled SPP1 molecule (figure 23-4B; 8). A left-handed G*35*P filament interacts with G*40*P-ATP and G*36*P. G*40*P-ATP bound at the ssDNA region free of G*36*P directs the assembly of DnaG and DNA PolIII (6, 79). The 3′-OH end of the annealed strand might act as a primer for initiating concatemeric DNA synthesis (sigma replication) on a supercoiled template. Furthermore, G*38*P bound to *ori*R might interact with the G*39*P–G*40*P–ATP complex to enhance the loading of the DNA helicase at the D-loop region. In the case of phage λ sigma replication, it was suggested that the λ-O protein bound to the λ-*ori* might create a physical barrier that permits only one round of unidirectional theta replication both in vivo and in vitro (11, 42). The broken fork could be rescued either by the concerted action of Red products (λ-α and λ-β proteins, "functional counterpart" of SPP1 G*35*P and G*34.1*P) or by the host recombination and repair machinery (see 8, 107, 112). This is consistent with the fact that the Red genes are not essential, though involved in λ replication (107, 112). Since λ and SPP1 sigma replication models are alike, we suggested that G*34.1*P–G*35*P (in the case of phage SPP1) and the Red and/or the host recombination system (in the case of phage λ; 107) promote the formation of new replication forks, to rescue the broken replication forks, by the invasion of duplex DNA by a ssDNA 3′ terminus on a supercoiled homolog. The invading 3′-OH end then might be used to prime sigma replication.

Morphogenesis

Bacteriophage SPP1 devotes more than 40% of its genetic information to the synthesis and assembly of structural components (figures 23.2, 23.3). The phage capsid and the tail are formed in independent assembly pathways. The sequence of reactions that yields the DNA-filled capsid was studied in enough detail to recommend SPP1 as one of the model systems for viral capsid assembly (figure 23-5). In contrast, the tail assembly remains to be characterized in detail. The DNA-filled capsid is composed of six viral proteins (G*6*P, G*7*P, G*12*P, G*13*P, G*15*P, and G*16*P) and three non-structural SPP1 proteins participate in its

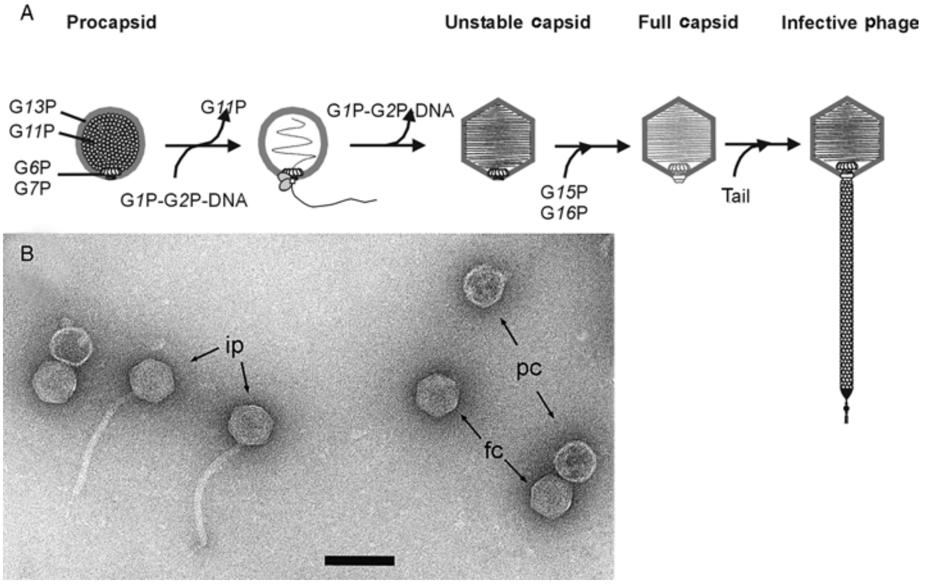


Figure 23-5 Morphogenetic pathway of bacteriophage SPP1. A: The first stable intermediate in phage assembly is the procapsid. It contains four proteins: major head protein, G13P (gray shell); scaffolding protein G11P (small circles that fill the inner space of the structure); and the portal protein, G6P, associated with the minor head protein, G7P, which are localized at the portal vertex of the procapsid (white structure). The terminase complex (G1P–G2P) is represented by gray ovals and a white square. After interaction of the terminase–DNA complex with the portal vertex the DNA (thin line) is pumped into the head, the scaffolding protein is released, and capsid expansion occurs. Termination of DNA packaging requires headful cleavage of the DNA concatemer and release of the terminase–DNA concatemer from the portal vertex. The DNA-filled capsid is stabilized by binding of G15P and G16P to the portal vertex. Attachment to the portal vertex of the phage tail that is assembled in an independent assembly pathway completes SPP1 assembly. B: The micrograph shows stable assembly intermediates of the SPP1 morphogenetic pathway and its final product, the virion. Abbreviation include: pc, procapsid; fc, DNA-filled capsid; ip, infective phage. The bar represents 100 nm. Negative staining with uranyl acetate. See thebacteriophages.org/frames_0230.htm for color version of this figure.

assembly (G1P, G2P, and G11P). The tail has at least eight different proteins as identified in SDS-PAGE patterns (51; Dröge and Stiege, personal communication). The role of host proteins in assembly has not been studied.

Procapsid Assembly

The first detectable intermediates of capsid assembly are spherically shaped procapsids with an outer diameter of approximately 55 nm. These are composed of the portal protein G6P (57.3 kDa), the minor capsid protein G7P (35.1 kDa), the scaffolding protein G11P (23.5 kDa), and the major capsid protein G13P (35.4 kDa) (12, 45; figure 23-5). The procapsid icosahedral lattice is formed by the major capsid protein G13P, displaying most probably T = 7 symmetry (45). The interior of the procapsid is "filled" with 100–180 copies of the scaffolding protein G11P. Coproduction of G11P and G13P yields procapsid-like structures both in *B. subtilis* infected with SPP1 DNA-packaging mutants or when genes *11* and *13* are coexpressed in a heterologous host (12, 45). However, no interaction is observed between the two proteins when produced independently and then mixed in vitro (12, 45). G13P alone polymerizes into curvilinear structures, most frequently with a spiral-like shape. Presence of G11P directs the polymerization of G13P to the correct geometry required for formation of closed protein lattices (12). The organization of G11P inside the procapsid is not known, but it was shown to adopt a polymorphic oligomeric state in solution. This organization is unique for the SPP1 scaffolding protein when compared with functional analogs from other species (12).

The procapsid-like structures composed of G11P and G13P are a mix of particles with the normal procapsid size (T = 7) and others with a smaller dimension (T = 3 or 4). None of these are competent for viral DNA encapsidation (45). The third component required for biological activity of the SPP1 procapsid is the portal protein G6P that defines a specialized vertex for chromosome packaging (45, 76). Portal proteins are cyclical oligomers with a turbine-like shape. Such structures are ubiquitous among tailed bacteriophages and herpesviruses (figure 23-6; 91, 93, 106, 126). Presence of G6P leads to the formation of procapsids with the normal size showing that the portal protein influences the copolymerization of G11P and G13P, preventing assembly of small procapsids (45). This effect

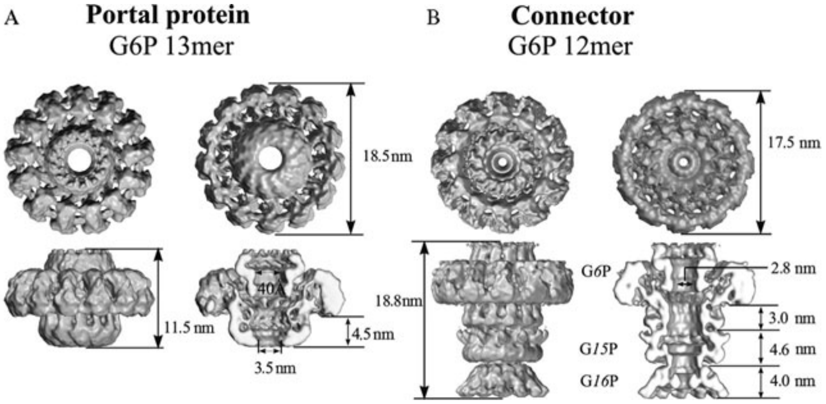


Figure 23-6 Three-dimensional structures of the bacteriophage SPP1 portal protein G6P and of the connector. The structures are based on data from electron micrographs of G6P oligomers and connectors imaged under the electron microscope and processed by single particle analysis (94). A: Different views of the G6P cyclical 13-mer and a cut-open view along the symmetry axis of the molecule (bottom right). In B: Equivalent views of the connector composed of stacked rings of G6P, G*15*P, and G*16*P. Each ring is composed of 12 subunits. The wide region of the connector formed by G6P is oriented toward the interior of the viral capsid while the bottom is exposed to the phage tail. The connector internal channel that can accommodate a DNA molecule is closed at the level of G*16*P (bottom ring) preventing exit of DNA from the phage capsid (Courtesy of E. V. Orlova). See thebacteriophages.org/frames_0230.htm for color version of this figure.

implies that the portal protein must be integrated in the procapsid structure at an early stage since the decision between formation of a T = 7 procapsid or of a structure with a smaller size is made after the first round of hexamers is assembled around the initial procapsid vertex (121). G6P is most likely a component of the initiator complex of SPP1 procapsid assembly, but it does not act as a nucleator of the reaction because the rate of assembly in vivo is independent of its presence (45). The organization of the initiator complex is not known. A stable interaction between G6P, G*11*P and/or G*13*P is detected only when the three proteins are coproduced, suggesting that the efficient incorporation of G6P into the procapsid requires a structural context created by G*11*P and G*13*P (45). Single amino acid substitutions in G6P that block its incorporation in the procapsid structure cluster in two distinct segments of the G6P primary structure (66). These mutations identify regions in the portal protein that are required either for G6P association into cyclical oligomers or for interaction with G*11*P and/or with G*13*P during procapsid assembly (66).

The portal protein G6P is present as a cyclical 12-mer in the SPP1 virion (76). However, G6P alone is organized as a cyclical oligomer with 13-fold symmetry in solution and in three-dimensional crystals (46, 68, 69). Such symmetry is an intrinsic property of the protein that is maintained upon dissociation–reassociation (68) and upon denaturation followed by refolding and reassociation (Jekow and Orlova, personal communication). The architecture of the G6P 13-mer (93; figure 23-6) is remarkably similar to that of portal proteins of other species in spite of there being no detectable relationship between their amino acid sequences (126). Lurz et al. (76) argued that the 13-mer is an on-pathway precursor of the phage portal oligomer and that G6P acquires a 12-mer organization in the initiation complex of procapsid assembly. The G6P forms competent for assembly of this complex would be gp6 open curvilinear oligomers (. . .8-, 9-, 10-, 11-, 12-mers) found in equilibrium with the closed 13-mer ring (127). The interaction with the other procapsid proteins, G*11*P and G*13*P, would impose an increased bend between gp6 subunits, locking them in a 12-mer ring (76). Under these conditions procapsid assembly, an irreversible process, displaces the equilibrium of G6P toward dissociation of the 13-mer ring, feeding additional subunits and curvilinear oligomers for procapsid assembly.

The major capsid protein, the scaffolding protein, and the portal protein are the common requirements to assemble procapsids competent for DNA packaging in all tailed bacteriophages (phage HK097 and possibly the Sfi21-like group of phages are exceptions; see 22, 62). These are also the only essential components for assembly of SPP1 biologically active procapsids (45). The fourth component of the SPP1 procapsid is G7P present in two to three copies per particle (12, 43, 115). In the absence of G7P, viable phages are still assembled in an in vitro DNA packaging assay. However, the procapsid biological activity is reduced 5 to 10-fold (45). G7P forms stable complexes with G6P in vivo and in vitro and is most likely localized in the portal vertex of the procapsid (115). Formation of this complex seems essential for G7P to accomplish its function, but the precise role of G7P in the SPP1 life cycle remains to be determined. The G6P–G7P interaction is disrupted during assembly of the virion, probably when DNA is translocated

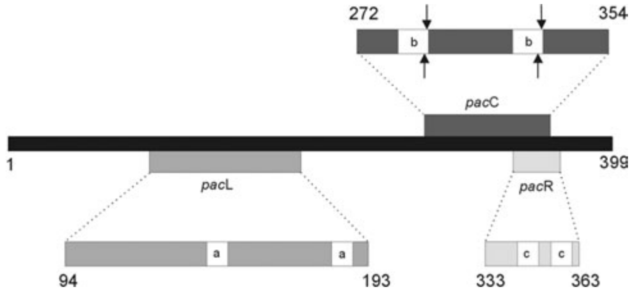


Figure 23-7 The *pac* region of bacteriophage SPP1. The black bar indicates 399 base pairs of the SPP1 sequence containing the packaging initiation region. The enlarged regions, *pac*L, and *pac*R, are recognized by the terminase small subunit (G*1*P) and are the nonencapsidated left end and the encapsidated right end, respectively, of the SPP1 sequence. *pac*C is the packaging processing site. The open boxes labeled a, b, and c indicate repeated sequences in the *pac* region. The arrows indicate the staggered nicks in the 10 bp repeat boxes within *pac*C introduced by the terminase large subunit (G*2*P). See thebacteriophages.org/frames_0230.htm for color version of this figure.

through the portal pore to the capsid interior (115). Proteins with these properties were not characterized in other virus systems. However, the finding of genes whose products share similarity with G*7*P in several bacteriophage species of Gram-positive bacteria (75, 115, 133 and references therein) indicates that this protein has a widespread role in a variety of phage systems.

DNA Packaging

SPP1 uses concatemers of its genome as the substrate for packaging by a headful mechanism (see chapter 6). Encapsidation is initiated by recognition and cleavage at the nucleotide sequence *pac* followed by unidirectional translocation of concatemer DNA to the interior of the preformed procapsid (16, 20, 29, 32, 59). When a threshold amount of DNA, representing about 104% of the SPP1 genome, has been packaged (headful) a sequence-independent DNA cleavage terminates the encapsidation cycle (89, 120). A second cycle of encapsidation initiates at the DNA extremity generated by this cut and sequential headful packaging events proceed along the DNA concatemer in a processive fashion (89, 119). The headful packaging mechanism generates a heterogeneous population of terminally redundant and partially circularly permuted DNA molecules as in the cases of P1 and P22 or of T4 whose DNA is totally permuted and terminally redundant (16, 17, 55).

The Terminase and Initiation of DNA Packaging

Initiation of phage DNA packaging involves the specific interaction of the procapsid with virus DNA. This process is mediated by viral nonstructural proteins, called terminases, which specifically recognize viral DNA and cleave it (reviewed in 16, 55, 90). The SPP1 terminase is composed of two subunits that achieve these two functions in the absence of other phage products (32, 59).

The SPP1 terminase enzyme hetero-oligomers are composed of decameric G*1*P and monomeric G*2*P (28, 30, 58, 59). The terminase small subunit, G*1*P, is a 183 amino acid long polypeptide that recognizes specifically the bipartite packaging initiation site [*pac*L (nonencapsidated DNA end or left end) and *pac*R (the end to be encapsidated or right end); figure 23-7]. G*1*P, upon interacting with *pac*L and *pac*R, forms a nucleoprotein structure that helps to position the terminase large subunit, G*2*P, at the packaging processing site, *pac*C (32, 37) and interacts with G*6*P (25). G*2*P, which is a 422 amino acid polypeptide, introduces two 2 bp staggered nicks at each of the 10 bp repeat, box b sites, found within *pac*C (28, 29, 30, 32, 37, 58, 59). G*1*P enhances approximately 4-fold the G*2*P ATPase activity, but exerts a negative effect on its endonuclease activity (59). Presence of G*6*P enhances further the G*2*P ATPase activity, probably because the proteins form a complex that becomes active for DNA translocation (25). The DNA end protected by the terminase (encapsidated end) is then translocated into the prohead in an ATP-driven reaction, whereas the nonencapsidated end (*pac*L) is degraded (25, 32, 59; figure 23-8).

The SPP1 DNA packaging reaction was reproduced in vitro by combination of extracts of *B. subtilis* cells infected with SPP1 mutants that are defective either in synthesis of one terminase subunit or in the assembly of the procapsid (44). Biologically active procapsids are present in the first extract (procapsid donor) and terminase in the second (terminase donor), respectively. SPP1 DNA is present in both extracts. Upon mixture of the two extracts all requirements for phage DNA packaging are present and infective virions assemble. A noteworthy result is that the DNA packaged originates exclusively from the labile terminase donor extract. Terminase thus appears to be already loaded in the substrate DNA when extracts are mixed while free active enzyme that could bind SPP1 DNA in the procapsid donor extract is not available. The terminase–DNA complex was proposed to be significantly more stable than free terminase and would be the active form for the packaging reaction under conditions where procapsid assembly and DNA encapsidation are uncoupled (44).

DNA Translocation

The SPP1 DNA–terminase complex docks at the portal vertex of the procapsid to initiate DNA packaging. This complex probably contains a terminase that is simultaneously bound to the SPP1 *pac* site and to the procapsid (figure 23-8). DNA translocation is then initiated

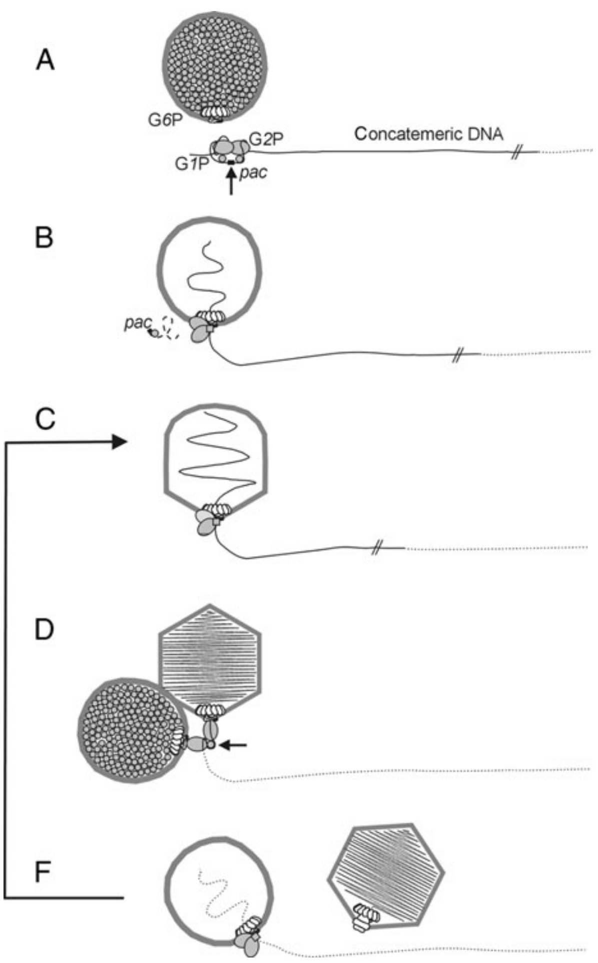


Figure 23-8 DNA packaging in bacteriophage SPP1. The SPP1 DNA packaging has been divided in five different steps. 1: The terminase complex (G*1*P–G*2*P, represented as gray ovals and smaller gray circles, respectively) is assembled at one *pac* site of the phage DNA concatemer and recognizes the portal protein region of the procapsid (representation of the individual components is as in figure 23-5). The arrow indicates the staggered nicks in *pac*C (see figure 23-7). 2: After DNA cutting at *pac* the encapsidated end (*pac*R) is translocated into the procapsid and the nonencapsidated end (*pac*L) is being degraded. 3: With translocation of the DNA the scaffolding protein is released and the capsid shell expands. 4: When the capsid is filled with DNA the G6P sensor will give the signal for cutting the DNA and transfer the terminase–DNA concatemer complex to another procapsid. 5: After the headful cut, DNA encapsidation starts again with a new procapsid. The DNA-filled capsid proceeds to the final steps of phage morphogenesis. See thebacteriophages.org/frames_0230.htm for color version of this figure.

with a concomitant exit of the scaffolding protein G*11*P from the procapsid interior, thereby creating empty space for DNA packing. The DNA packaging motor is likely composed of the terminase, whose ATPase activity fuels DNA translocation, and of the portal protein through whose central channel the DNA is pumped. A significant number of point mutations in G*6*P block or reduce the efficiency of SPP1 DNA packaging, confirming its central role in this process (66).

The molecular mechanism of how viral DNA is mechanically translocated to the capsid interior is not understood in any phage system. It is not clear whether the DNA transport is a linear translocation or whether it involves a relative rotation between portal protein and DNA. Dube et al. (46) proposed a model for DNA translocation that exploits the symmetry mismatch between the helical symmetry of DNA and the symmetry of the portal protein. In this model the portal protein would rotate relative to the capsid and to the DNA helix, as initially proposed by Hendrix (61), to keep a constant positioning of the portal subunits relative to the DNA being packaged. This constant topology would be necessary for the portal subunits to carry out mechanical translocation of DNA in 2 bp steps powered by the terminase ATPase activity. The DNA would be translated to the capsid interior without significant rotations relative to the capsid. The model was formulated for a 13-mer symmetric portal protein but applies also to a 12-mer oligomer (46). Alternatively, G*2*P bound to DNA pulls it unidirectionally toward the G*1*P–G*6*P complex in an ATP-dependent process. The translocating G*2*P hydrolyzes ATP during its translocation along the linear lattice of double-stranded DNA. The translocated DNA is then passed through the central hole of the G*6*P turbine. This translocating mechanism resembles the inchworm model initially described for single-strand translocases or DNA helicases (see 108). Once the capsid is full, by an uncharacterized step, G*6*P might transfer the headful signal to G*2*P. Translocase activity would be subsequently attenuated, and a weak nuclease activity would be activated (25; see also below).

During DNA packaging the subunits of the major capsid protein, G*13*P, undergo a conformational change which results in both expansion of the shell to the mature size and an increased angularity of the structure (figure 23-5 and 23-8; 12, 45).

Packaging Termination: The Headful Packaging Mechanism

When the head is full, the DNA inside the shell is separated from the DNA remaining outside by a headful cleavage. This sequence-independent and also imprecise cut (65, 119) is determined by the amount of DNA present inside the capsid. *Siz* mutations, which lead to undersizing of the DNA packaged, were only identified in the gene coding for the SPP1 portal protein G*6*P (120; see 27 for a related phenotype in bacteriophage P22). The most severe of these mutations are associated with deletions of nonessential regions of the SPP1 genome. The deletion was required to ensure presence of the complete set of phage essential genes in the shorter packaged chromosome.

Combination of *siz* mutations leads to encapsidation of DNA molecules significantly smaller than SPP1 mature chromosomes of single *siz* mutants (118). Additionally, the efficiency of DNA packaging is reduced leading to the suggestion that a trigger for headful cleavage could be the incapacity of the packaging machinery to encapsidate further DNA into the capsid (65, 118). Comparison between the structures of the SPP1 wild-type protein and a G*6*P form carrying a *siz* mutation showed significant differences in the crown. This crown is a flexible domain that delimits the central portal channel at the wider region of the G*6*P structure (figure 23-6; 93, 94). Its orientation to the capsid interior provides an ideal topology to act as a sensor of the mass of packaged DNA. The increasing pressure exerted by the DNA on the sensor domain was proposed to trigger a conformational change in this domain leading to immobilization of DNA inside the portal channel or to a stopping of the packaging machinery activity (93). This signal could alter the stoichiometry of the G*6*P:G*1*P:G*2*P translocase complex. An imbalance in the G*6*P:G*1*P:G*2*P complex, perhaps a low amount of G*1*P or an excess of G*6*P, "induces" G*2*P to introduce a nonspecific cleavage (headful cut), separating the packaged DNA from the concatemer (25, 59). The headful cleavage generates a new DNA end that serves as the starting point for a subsequent round of DNA packaging (32, 37, 118–120).

Stabilization of the Packaged DNA

After the terminase departs, the portal pore is plugged by the head completion proteins, G*15*P and G*16*P, to avoid release of the packaged DNA (12, 76, 94). A mutation in G*6*P was shown to block specifically this step (see note added in proof on page 345). The structure assembled at the portal vertex (connector) comprises stacked rings of G*6*P, G*15*P, and G*16*P subunits crossed by a central channel that is closed at the level of G*16*P (figure 23-6; 76, 94). In phages whose capsid was disrupted, the connector remains attached to the tail structure through the G*16*P ring region and the complex has been shown to protect a DNA fragment of about 20 nm (119). This size fits particularly well to the connector height (figure 23-6). The DNA extremity fixed to the connector was the last one to be packaged and is the first to exit the virion at the beginning of infection (119). The connector thus acts as a valve that retains DNA inside the capsid until a signal for release is triggered by interaction of SPP1 with its receptor at the *B. subtilis* cell surface. This signal must be transmitted through the complete tail structure from the tail fiber to the connector (figure 23-1).

Tail Assembly

The SPP1 tail is assembled in a morphogenetic pathway independent from capsid assembly. Its flexible helical region is composed of G*17.1*P and of a lower amount of G*17.1**P (figure 23-1). The two proteins have the same N-terminus but G*17.1**P has a higher molecular weight than G*17.1*P (48; A. Dröge and F. Weise, personal communication). No information is available about the components of the SPP1 tail fiber. Attachment of the tail to the connector, most probably to G*16*P (76), yields the SPP1 infective phage.

Lysis

Lysis of *B. subtilis* cells infected at 37 °C with SPP1 starts around 30 minutes post-infection. The mechanisms of lysis and how its timing is controlled were not studied. Two proteins that share identity with holins (G*26*P) and *N*-acetylmuramoyl-1-alanine amidases (G*25*P) are encoded by the SPP1 genome (3). These are the most likely effectors of lysis (see chapter 10).

SPP1 as a Genetic Tool

Transduction

Bacteriophage SPP1 mediates generalized transduction of host chromosomal markers (134) and plasmid transduction (26). The substrate for packaging of transducing DNA can be either chromosomal DNA of the host or concatemeric plasmid DNA. Both substrates are packaged by an identical mechanism.

SPP1 transducing particles transfer approximately 1% of the host *B. subtilis* chromosome from a donor to a receptor strain with an efficiency of 10^{-7} to 10^{-8} transductants per viable phage (134). This feature made SPP1 a tool widely used for genetic mapping of the *B. subtilis* chromosome. The transducing particles are indistinguishable from SPP1 virions with the exception that they carry a DNA molecule of bacterial origin with a size similar to that of the SPP1 chromosome (39, 40).

Plasmids, which are usually smaller than a bacteriophage genome, can be encapsidated into phage procapsids to produce plasmid transducing particles (74). The encapsidation of linear plasmid concatemers has been detected in all bacteria–phage systems tested so far (2, 24, 38, 70, 92, 105). Plasmid transduction frequencies are generally low, about 10^{-5} to 10^{-7} per active phage particle. However, when homology is provided between the phage genome and the plasmid, transduction frequency increases up to 10^5-fold (transduction facilitation effect) (2, 38, 92, 105). With *B. subtilis* phage SPP1, homology of as little as 47 bp provided a maximal facilitation effect (2). Beyond this observation, however, the molecular events that trigger the synthesis of packagable concatemeric plasmid DNA are not well understood. In bacteria deficient in ExoV (also termed RecBCD or AddAB activity) the accumulation of high molecular weight (hmw) linear head-to-tail plasmid concatemers was observed (34, 130).

SPP1 infection of *B. subtilis* cells bearing plasmids induces the synthesis of concatemeric plasmid DNA, and such hmw DNA synthesis is independent of plasmid-encoded replication functions (19). Accumulation of hmw plasmid DNA, starting simultaneously with the onset of SPP1 sigma replication (24), was dependent on phage-encoded function(s) but independent of host-encoded primosomal functions (see 131).

The molecular events that trigger the synthesis of packagable head-to-tail chimeric SPP1:plasmid concatemeric SPP1 DNA might partially follow the mechanism used by SPP1 RDR (see above). Upon SPP1 infection, a phage-encoded product might directly or indirectly inactivate the AddAB enzyme with the subsequent accumulation of plasmid stalled replication forks. The repair of this stalled replication fork requires the SPP1-encoded G*35*P and G*34.1*P recombination proteins to create a D-loop (a single-stranded 3′-OH-end of phage DNA or plasmid DNA invading a homologous region of the dsDNA plasmid molecule) (R. Missich and S. Ayora, personal communication). The synapsis between a phage and a plasmid molecule would lead to the generation of a phage:plasmid DNA chimera carrying the *pac* site. The terminase enzyme recognizes and cleaves the *pac* site present in the chimera to initiate unidirectional and processive "headful" packaging (32, 70). In the absence of apparent homology between the plasmid and the phage, however, the hmw DNA could be packaged at a very low frequency into empty procapsids by an uncharacterized mechanism.

Transfection

Exposure of competent bacterial cells to bacteriophage DNA, leading to the production of intact viruses in such cells, was first observed in *B. subtilis* and its phage SP50 and called "transfection" (a recombinant word from transformation and infection) (52). A peculiar feature of transfection in *B. subtilis* is its dose-response, that is the relationship between the concentration of input phage DNA used in the transfection experiment and the number of cells yielding a phage burst (123). Whereas a linear relationship is observed in infection, transfection is always characterized by a quadratic or even higher order dependence on the DNA concentration. This dose-response was observed over a wide range of DNA concentrations. It is assumed that this relationship reflects different modes of entry of phage DNA molecules into the cell: In infection, the mature phage chromosome is transported into the interior of the cell by the phage ejection mechanism. This dsDNA molecule that contains the complete viral genome will immediately enter the replication cycle. Transfection, however, follows the entry mode of bacterial or plasmid transformation in *B. subtilis*. Here double-stranded transforming DNA is fragmented following attachment to competent cells (47). Of each fragment only one strand is taken up by the cells, the other being degraded (96). Most likely, such uptake is polar with respect to the 5′ to 3′ orientation of DNA, as in *S. pneumoniae* (82). In such a scenario only the uptake of at least two phage DNA molecules in independent entry events will provide a transfected cell with an ensemble of phage DNA fragments derived from strands with different strand polarity. Alignment of such fragments by annealing within regions of sequence complementarity, followed by gap filling DNA synthesis, would generate a replicative molecule. Fragmentation of transfecting DNA as the cause of a multi-powered dose-response was first suggested for transfection with *B. subtilis* phage SP82 (57). The processing of transfected DNA as described is compatible with the observation that, in contrast to transformation, transfection is sensitive to restriction (122, 124). Phage DNA loses transfecting activity following degradation by a restriction endonuclease since, in spite of uptake of such DNA, no contiguous DNA molecule can be formed intracellularly. In line with the model of DNA processing discussed, however, combinations of two endonuclease digests have transfection activity, provided the restriction fragments generated have large overlaps (64, 120).

If transfection is performed by mixtures of two genetically distinguishable SPP1 DNAs differing in two genes, where recombinants can be scored in the progeny, recombination is significantly enhanced over the level observed in conventional phage crosses (110). This would be anticipated with the patching-up model described.

The postulated uptake mechanism in transfection also explains the low efficiency of transfection in comparison with infection. Apparently many cells take up phage DNA without being able to process it to replicative progeny. Such abortive transfection becomes obvious under marker rescue conditions, when competent cells were infected with conditionally lethal mutant phages simultaneously or prior to exposure to wild-type phage DNA. Under such conditions, in which the wild-type allele of the DNA can substitute for the mutant allele of the "helper" phage, the efficiency of transfection is dramatically enhanced, and the dose-response of transfection becomes linear. Also, fragmented wild-type DNA, otherwise inactive, will serve under these conditions. Such marker rescue experiments using a wide spectrum of mutant helper phages and individual SPP1 DNA restriction fragments have led to an unambiguous assignment of mutations to restriction fragments, thus providing a coarse correlation between the genetic and physical maps of the phage (13).

Purified complementary single strands of SPP1 DNA (neither of which is alone active in transfection) could be annealed to produce transfecting molecules. When such annealing was performed with single strands derived from genetically different and distinguishable SPP1 phages, one could generate in each combination two types of heteroduplex molecule, in which the alleles from each "parental" molecule were carried either on the "heavy" or "light" strand. Performing single-burst analyses of transfections with such heteroduplex DNA in combinations of wild-type DNA and the DNAs of a large number of mutants, we observed pure bursts of either wild-type, or mutant, or mixed bursts (109, 125). It was intriguing that in a given pair of heteroduplexes the frequencies of pure wild-type and pure mutant bursts were not identical. There was a strong bias for one of the two phenotypes, which was strand-specific rather than allele-specific. This observation suggests strand discrimination before the onset of phage replication. At the time when these results were obtained, neither the intricacies of DNA processing in transfection nor the architecture of SPP1 molecules were known. Therefore we assumed that a heteroduplex molecule would be taken up as such and that the mismatched site would induce repair of the mismatch similar to a gene conversion process ("conversion model"). At the present stage this interpretation must be revised on the basis of information available on DNA uptake. It is conceivable that in the uptake of heteroduplex molecules one strand is lost with a higher frequency than the other and that the strand preferentially taken up serves as a template in the patch-up synthesis described before ("strand-selection model"). However, neither the "conversion model" nor the "strand-selection model" can explain the strong strand bias observed in our experiments. Bacterial and/or phage mutations deficient in mismatch repair, if available, would be helpful in assessing the contribution of each of the two processes to the known observations.

SPP1 as a Cloning Vector

SPP1 phages have also been useful as cloning vectors. In these constructs part of the dispensable region of *Eco*RI fragment 1 of SPP1 DNA was eliminated and a unique restriction site, either *Bam*HI (SPP1v; 60) or *Pst*I (SPP1vic; 14), engineered into the fragment. Restriction fragments of heterologous DNA of up to 4 kb could be maintained and expressed in these vectors.

Evolution

The SPP1-Like Phages

The *B. subtilis* phages ρ15, SF6, and 41c (phage 22a is a different isolate of the same virus), isolated in different geographical regions (67, 114), are highly related to bacteriophage SPP1 by serological, biological, biochemical, and genome sequence similarity criteria (77, 103, 104; see "Adsorption and Infection"). DNA heteroduplex analysis showed that all phage DNAs shared extensive regions of homology interrupted by short nonhybridizing regions. Crosses between the phages yielded viable hybrid progeny, showing that this group of viruses share a common genetic pool (101; M.A. Santos, personal communication).

SPP1, the Prototype of a Wide Group of Gram-Positive Phages

SPP1 proteins exhibit similarity to the amino acid sequence of a variety of phage proteins that infect other Gram-positive bacteria of genera such as *Streptococcus*, *Lactobacillus*, *Listeria*, and *Staphylococcus* (3, 21, 32, 73, 95, 113, 115; chapter 4). Detectable similarity between SPP1 and each of those phages is limited to a few proteins. Biochemical and/or functional studies in the SPP1 system showed that these correspond frequently to groups of proteins that interact during SPP1 multiplication (e.g., the two terminase subunits G1P–G2P (58); G6P–G7P (45, 115); the connector proteins G15P–G16P and phage tail proteins (76, Dröge, personal communication). This observation suggests that the homologous proteins from other phage species have a function similar to SPP1 proteins. Constraints generated by the interaction between proteins thus appear to be an important factor that might slow down their sequence divergence due to coevolutionary requirements. This feature would facilitate identification of relationships between phages. Functional, biochemical, and structural information is a major requirement for understanding the forces driving virus evolution. The wealth of genetic and biochemical data on the molecular mechanisms that support the different aspects of the SPP1 life cycle strongly recommend it as the prototype for the group of phages described above that includes species of medical and economic interest.

Detectable relatedness of SPP1 proteins at the amino acid sequence level is essentially found with *Siphoviridae* phages infecting bacteria closely related phylogenetically but that can occupy different ecological niches. The evolution of these phages of Gram-positive bacteria seems therefore to be mostly shaped by the evolutionary history of their hosts. Host specificity appears as an effective barrier that avoids frequent horizontal transfer between virus infecting distantly related bacteria, even though some exceptions were identified (23). Sequence similarity is eroded by the long-distance evolution of phages, but analysis of their gene order and of the molecular strategies used for viral multiplication provides evidence of ancient evolutionary relatedness between SPP1 and the other tailed bacteriophages (see below; 23).

Gene Order and Biochemical Strategies: Long Term Evolutionary Conservation

The SPP1 genomic organization in the packaging and morphogenesis operons and in the initiation-of-replication region resembles those in coliphages λ and P22, but no or very little conservation was detected at the level of the deduced amino acid sequence (32, 36, 48, 95).

Conservation of the genomic organization in the packaging and head morphogenesis operons in the absence of amino acid sequence similarity between analogous proteins can be the result of two different types of evolutionary mechanisms. The amino acid differences between isofunctional proteins of phylogenetically distant phages infecting Gram-positive and Gram-negative hosts can be attributed if the principle of neutral evolution applies to a constant rate of amino acid substitution after divergence from a common ancestor. The order of the genes would be kept and reflect the arrangement present in a common phage ancestor. This hypothesis would imply a coevolutionary process that, in spite of drastic modifications in the polypeptides, would still allow productive interaction between the various morphogenetic proteins. Alternatively, we can assume the existence of different ancestors, with morphogenetic genes brought together by convergent evolution. This would readily explain the nonrelatedness among primary structures of the proteins but also would imply that: (i) numerous proteins evolved independently to accomplish the same function in the construction of a complex macromolecular structure using a remarkably similar assembly strategy, and (ii) the genes independently became clustered in a conserved order in distant lineages.

Independent of the evolutionary mechanism that preserved or, alternatively, brought together morphogenetic genes, the conservation of the gene arrangement indicates that it is an advantageous property for the phage. The driving force for this process could be the selective pressure to cluster cistrons whose products intensively interact (111) such that a genomic module would correspond to a functional unit (the replication apparatus, the packaging machinery, etc.). Fisher (53) has predicted that if some allele combination is superior to other combinations, such linkage can be transmitted intact, suffering only minimal disruption by recombination.

Note Added in Proof

The *Bacillus subtilis* receptor for bacteriophage SPP1 was recently shown to be the transmembrane protein YueB. (São-José, C., C. Baptista, and M.A. Santos. 2004. *Bacillus subtilis* operon encoding a membrane receptor for bacteriophage SPP1. *J. Bacteriol.* 186:8337–8346.).

Mutations in the SPP1 portal protein G6P that block specifically plugging of the portal pore after termination of DNA packaging were recently described (Isidro, A., A. O. Henriques, and P. Tavares. 2004. The portal protein plays essential roles at different steps of the SPP1 DNA packaging process. Virology 322:253–263.

Acknowledgments

We acknowledge past and present researchers on the biology of SPP1 bacteriophage for all their contributions to our current knowledge of this biological system. We thank Mário A. Santos for comments on parts of the manuscript. We thank particularly Brigitte Pawlek, Bernhard Behrens, Gerhild Lüder, and Asita Stiege for their continued and dedicated technical assistance. We are indebted to Silvia Ayora, Ana Camacho, Anja Dröge, Petra Jekow, Gerhild Lüder, María Martinez, Pablo Mesa, Riccardo Missich, Elena V. Orlova, Mário A. Santos, Asita Stiege, and Frank Weise for communication of unpublished results and to Elena V. Orlova for figure 23-6. Research in this system was partially supported by grants BMC2003-00150 from MCT-DGI and QLK2-CT-2000-00634 from the EU to J. C. A., and ATIPE from the CNRS and a subvention from FRM to P. T.

References

1. Ackermann, H.-W. 1999. Tailed bacteriophages: the order *Caudovirales*. Adv. Virus Res. 51:135–201.
2. Alonso, J. C., G. Lüder, and T. A. Trautner. 1986. Requirements for the formation of plasmid-transducing particles of *Bacillus subtilis* bacteriophage SPP1. EMBO J. 5:3723–3728.
3. Alonso, J. C., G. Lüder, A. C. Stiege, S. Chai, F. Weise, and T. A. Trautner. 1997. Analysis of the complete nucleotide sequence and functional organisation of *Bacillus subtilis* bacteriophage SPP1. Gene 204:201–212.
4. Amann, E. P., and J. N. Reeve. 1981. Expression of lambda/SPP1 hybrid phages in *E. coli* minicells. Mol. Gen. Genet. 182:299–303.
5. Amann, E. P., J. N. Reeve, G. Morelli, B. Behrens, and T. A. Trautner. 1981. Cloning and expression of the *Bacillus subtilis* phage SPP1 in *Escherichia coli*. I. Construction and characterization of lambda/SPP1 hybrids. Mol. Gen. Genet. 182:293–298.
6. Ayora, S., U. Langer, and J. C. Alonso. 1998. *Bacillus subtilis* DnaG primase stabilises the SPP1 G*40*P helicase–ssDNA complex. FEBS Lett. 439:59–62.
7. Ayora, S., A. Stasiak, and J. C. Alonso. 1999. The *Bacillus subtilis* bacteriophage SPP1 G*39*P delivers and activates the G*40*P DNA helicase upon interacting with the G*38*P-bound replication origin. J. Mol. Biol. 288:71–85.
8. Ayora, S., R. Missich, P. Mesa, R. Lurz, X. Yang, E. H. Egelman, and J. C. Alonso. 2002. The replication protein G*35*P of *Bacillus subtilis* bacteriophage SPP1 mediates homologous DNA pairing. J. Biol. Chem. 277:35969–35979.
9. Bailey, S., S. E. Sedelnikova, P. Mesa, S. Ayora, J. P. Waltho, A. E. Ashcroft, A. J. Baron, J. C. Alonso, and J. B. Rafferty. 2003. Structural analysis of *Bacillus subtilis* SPP1 phage helicase loader protein G*39*P. J. Biol. Chem. 278:15304–15312.
10. Bárcena, M., C. San Martin, F. Weise, S. Ayora, J. C. Alonso, and J. M. Carazo. 1998. Polymorphic quaternary organization of the *Bacillus subtilis* bacteriophage SPP1 replicative helicase. J. Mol. Biol. 283:809–819.
11. Bastia, D., and N. Sueoka. 1975. Studies on the late replication of phage lambda: rolling circle replication of the wild

type and a partially suppressed strain, Oam29 Pam80. J. Mol. Biol. 98:305–320.

12. Becker, B., N. de la Fuente, M. Gassel, D. Günther, P. Tavares, R. Lurz, T. A. Trautner, and J. C. Alonso. 1997. Head morphogenesis genes of the *Bacillus subtilis* bacteriophage SPP1. J. Mol. Biol. 268:822–839.
13. Behrens, B., G. Lüder, M. Behnke, and T. A. Trautner. 1979. The genome of *B. subtilis* phage SPP1: physical arrangement of phage genes. Mol. Gen. Genet. 175:351–357.
14. Behrens, B., B. Pawlek, G. Morelli, and T. A. Trautner. 1983. Restriction and modification in *B. subtilis*: construction of hybrid lambda and SPP1 phages containing a DNA methyltransferase gene from *B. subtilis* phage SPR. Mol. Gen. Genet. 189:10–16.
15. Biswal, N., A. K. Kleinschmidt, H. C. Spatz, and T. A. Trautner. 1967. Physical properties of the DNA of bacteriophage SP 50. Mol. Gen. Genet. 100:39–55.
16. Black, L. W. 1989. DNA packaging in dsDNA bacteriophages. Annu. Rev. Microbiol. 43:267–292.
17. Black, L. W. 1995. DNA packaging and cutting by phage terminases: control in phage T4 by a synaptic mechanism. BioEssays 17:1025–1030.
18. Bradley, D. E. 1967. Ultrastructure of bacteriophages and bacteriocins. Bacteriol. Rev. 31:230–314.
19. Bravo, A., and J. C. Alonso. 1990. The generation of concatemeric plasmid DNA in *Bacillus subtilis* as a consequence of bacteriophage SPP1 infection. Nucleic Acids Res. 18:4651–4657.
20. Bravo, A., J. C. Alonso, and T. A. Trautner. 1990. Functional analysis of the *Bacillus subtilis* bacteriophage SPP1 *pac* site. Nucleic Acids Res. 18:2881–2886.
21. Brøndsted, L., S. Østergaard, M. Pedersen, K. Hammer, and F. K. Vogensen. 2001. Analysis of the complete DNA sequence of the temperate bacteriophage TP901-1: evolution, structure, and genome organization of lactococcal bacteriophages. Virology 283:93–109.
22. Brüssow, H., and F. Desiere 2001. Comparative phage genomics and the evolution of *Siphoviridae*: insights from dairy phages. Mol. Microbiol. 39:213–222.
23. Brüssow, H., and R. W. Hendrix. 2002. Phage genomics: small is beautiful. Cell 108:13–16.
24. Burger, K. J., and T. A. Trautner. 1978. Specific labelling of replicating SPP1 DNA: analysis of viral DNA synthesis and identification of phage DNA-genes. Mol. Gen. Genet. 166:277–285.
25. Camacho, A. G., A. Gual, R. Lurz, P. Tavares, and J. C. Alonso. 2003. *Bacillus subtilis* bacteriophage SPP1 DNA packaging motor requires terminase and portal proteins. J. Biol. Chem. 278: 23251–23259.
26. Canosi, U., G. Lüder, and T. A. Trautner. 1982. SPP1-mediated plasmid transduction. J. Virol. 44:431–436.
27. Casjens, S., E. Wyckoff, M. Hayden, L. Sampson, K. Eppler, S. Randall, E. T. Moreno, and P. Serwer. 1992. Bacteriophage P22 portal protein is part of the gauge that regulates packing density of intravirion DNA. J. Mol. Biol. 224:1055–1074.
28. Chai, S., and J. C. Alonso. 1996. Distamycin-induced inhibition of formation of a nucleoprotein complex between the terminase small subunit G1P and the non-encapsidated end (*pac*L site) of *Bacillus subtilis* bacteriophage SPP1. Nucleic Acids Res. 24:282–288.
29. Chai, S., V. Kruft, and J. C. Alonso. 1994. Analysis of the *Bacillus subtilis* bacteriophages SPP1 and SF6 gene *1* product: a protein involved in the initiation of headful packaging. Virology 202:930–939.
30. Chai, S., R. Lurz, and J. C. Alonso. 1995. The small subunit of the terminase enzyme of *Bacillus subtilis* bacteriophage SPP1 forms a specialized nucleoprotein complex with the packaging initiation region. J. Mol. Biol. 252:386–398.
31. Chai, S., U. Szepan, and J. C. Alonso. 1997. *Bacillus subtilis* bacteriophage SPP1 terminase has a dual activity: it is required for the packaging initiation and represses its own synthesis. Gene 184:251–256.
32. Chai, S., A. Bravo, G. Lüder, A. Nedlin, T. A. Trautner, and J. C. Alonso. 1992. Molecular analysis of the *Bacillus subtilis* bacteriophage SPP1 region encompassing genes *1* to *6*. The products of gene *1* and gene *2* are required for *pac* cleavage. J. Mol. Biol. 224:87–102.
33. Chai, S., U. Szepan, G. Lüder, T. A. Trautner, and J. C. Alonso. 1993. Sequence analysis of the left end of the *Bacillus subtilis* bacteriophage SPP1 genome. Gene 129:41–49.
34. Cohen, A., and A. J. Clark. 1986. Synthesis of linear plasmid multimers in *Escherichia coli* K-12. J. Bacteriol. 167:327–335.
35. Cox, M. M. 2001. Historical overview: searching for replication help in all of the *rec* places. Proc. Natl. Acad. Sci. USA 98:8173–8180.
36. Daniels, D. L., J. L. Schroeders, W. Szybalski, F. Sanger, A. R. Coulson, G. F. Hong, D. F. Hill, G. B. Petersen, and R. F. Blattner. 1983. Complete annotated lambda sequence, pp. 519–677. *In* R. W. Hendrix, J. W. Roberts, F. W. Stahl, and R. A. Weisberg (eds.) *Lambda II*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
37. Deichelbohrer, I., W. Messer, and T. A. Trautner. 1982. Genome of *Bacillus subtilis* bacteriophage SPP1: structure and nucleotide sequence of *pac*, the origin of DNA packaging. J. Virol. 42:83–90.
38. Deichelbohrer, I., J. C. Alonso, G. Lüder, and T. A. Trautner. 1985. Plasmid transduction by *Bacillus subtilis* bacteriophage SPP1: effects of DNA homology between plasmid and bacteriophages. J. Bacteriol. 162:1238–1242.
39. de Lencastre, H., and L. J. Archer. 1980. Characterization of bacteriophage SPP1 transducing particles. J. Gen. Microbiol. 117:347–355.
40. de Lencastre, H., and L. J. Archer. 1981. Molecular origin of transducing DNA in bacteriophage SPP1. J. Gen. Microbiol. 122:345–349.
41. Desmyter, A., J. N. Reeve, G. Morelli, and T. A. Trautner. 1985. Inversion and deletion mutant in *Bacillus subtilis* bacteriophage SPP1 as a consequence of cloning. Mol. Gen. Genet. 198:537.
42. Dodson, M., H. Echols, S. Wickner, C. Alfano, K. Mensa-Wilmot, B. Gomes, J. LeBowitz, J. D. Roberts, and R. McMacken. 1986. Specialized nucleoprotein structures at the origin of replication of bacteriophage λ: localized unwinding of duplex DNA by a six-protein reaction. Proc. Natl. Acad. Sci. USA 83:7638–7642.
43. Dröge, A. 1998. *Capsidmorphogenese des Bakteriophagen SPP1*. Doctoral thesis, Technische Universität Berlin, Germany.
44. Dröge, A., and P. Tavares. 2000. In vitro packaging of DNA of the *Bacillus subtilis* bacteriophage SPP1. J. Mol. Biol. 296:103–115.

45. Dröge, A., M. A. Santos, A. Stiege, J. C. Alonso, R. Lurz, T. A. Trautner, and P. Tavares. 2000. Shape and DNA packaging activity of bacteriophage SPP1 procapsid: protein components and interactions during assembly. J. Mol. Biol. 296:117–132.
46. Dube, P., P. Tavares, R. Lurz, and M. van Heel. 1993. Bacteriophage SPP1 portal protein: a DNA pump with 13-fold symmetry. EMBO J. 12:1303–1309.
47. Dubnau, D. 1976. Genetic transformation of *Bacillus subtilis* a review with emphasis on the recombination mechanism, pp. 14–27. *In* D. Schlesinger (ed.) *Microbiology 1976.* American Society for Microbiology, Washington, D.C.
48. Eppler, K., E. Wyckoff, J. Goates, R. Parr, and S. Casjens. 1991. Nucleotide sequence of the bacteriophage P22 genes required for DNA packaging. Virology 183:519–538.
49. Esche, H. 1975. Gene expression of bacteriophage SPP1. II. Regulatory aspects. Mol. Gen. Genet. 142:57–66.
50. Esche, H., and H. Spatz. 1973. Asymmetric transcription of SPP1 in vivo. Mol. Gen. Genet. 124:57–63.
51. Esche, H., M. Schweiger, and T. A. Trautner. 1975. Gene expression of bacteriophage SPP1. I. Phage directed protein synthesis. Mol. Gen. Genet. 142:45–55.
52. Földes, J., and T. A. Trautner. 1964. Infectious DNA from a newly isolated *B. subtilis* phage. Z. Vererbungsl. 95:57–65.
53. Fisher, R. 1930. *In* The Genetical Theory of Natural Selection. Clarendon Press, Oxford.
54. Formosa, T., and B. M. Alberts. 1986. DNA synthesis dependent on genetic recombination: characterization of a reaction catalyzed by purified bacteriophage T4 proteins. Cell 47:793–806.
55. Fujisawa, H., and M. Morita. 1997. Phage DNA packaging. Genes Cells 2:537–545.
56. Ganesan, A. T., J. J. Andersen, J. Luh, and M. Effron. 1976. DNA metabolism, pp. 319–325, *In* D. Schlessinger (ed.) *Bacillus subtilis* and its Phage SPP1. *Microbiology 1976*. American Society of Microbiology, Washington, D.C.
57. Green, D. M. 1966. Intracellular inactivation of infective SP82 bacteriophage DNA. J. Mol. Biol. 22:1–13.
58. Gual, A., and J. C. Alonso. 1998. Characterization of the small subunit of the terminase enzyme of the *Bacillus subtilis* bacteriophage SPP1. Virology 242:279–287.
59. Gual, A., A. G. Camacho, and J .C. Alonso. 2000. Functional analysis of the terminase large subunit, G2P, of *Bacillus subtilis* bacteriophage SPP1. J. Biol. Chem. 275:35311–35319.
60. Heilmann, H., and J. N. Reeve. 1982. Construction and use of SPP1v, a viral cloning vector for *B. subtilis.* Gene 17:91–100.
61. Hendrix, R. W. 1978. Symmetry mismatch and DNA packaging in large bacteriophages. Proc. Natl. Acad. Sci. USA 75:4779–4783.
62. Hendrix, R. W., and R. L. Duda. 1998. Bacteriophage HK97 head assembly: a protein ballet. Adv. Virus Res. 50:235–288.
63. Helmann, J. D. 1995. Compilation and analysis of *Bacillus subtilis* σ^A-dependent promoter sequences: evidence for extended contact between RNA polymerase and upstream promoter DNA. Nucleic Acids Res. 23: 2351–2360.
64. Humphreys, G. O., and T. A. Trautner. 1981. Structure of *Bacillus subtilis* bacteriophage SPP1 DNA in relation to its transfection activity. J Virol. 37:574–579.
65. Humphreys, G. O., and T. A. Trautner. 1981. Maturation of bacteriophage SPP1 DNA: limited precision in the sizing of mature bacteriophage genomes. J. Virol. 37:832–835.
66. Isidro, A., M. A. Santos, A. O. Henriques, and P. Tavares. 2004. The high-resolution functional map of bacteriophage SPP1 portal protein. Mol. Microbiol. 51:949–962.
67. Jacobson, E. D., and O. E. Landman. 1975. Interaction of protoplasts, L forms, and bacilli of *Bacillus subtilis* with 12 strains of bacteriophage. J. Bacteriol. 124:445–448.
68. Jekow, P., J. Behlke, W. Tichelaar, R. Lurz, M. Regalla, W. Hinrichs, and P. Tavares. 1999. Role of the ionic environment on the assembly of bacteriophage SPP1 portal protein. Eur. J. Biochem. 264:724–735.
69. Jekow, P., S. Schaper, D. Günther, P. Tavares, and W. Hinrichs. 1998. Crystallization and preliminary X-ray crystallographic studies of the 13-fold symmetric portal protein of bacteriophage SPP1. Acta Crystallogr. D54:1008–1011.
70. Kreuzer, K.N., W. Y. Yap, A. E. Menkens, and H. W. Engman. 1988. Recombination-dependent replication of plasmids during bacteriophage T4 replication. J. Biol. Chem. 263:11366–11373.
71. Kunst, F., N. Ogasawara, I. Moszer, A. M. Albertini, G. Alloni, V. Azevedo, M. G. Bertero, P. Bessieres, A. Bolotin, S. Borchert, R. Borriss, L. Boursier, A. Brans, M. Braun, S. C. Brignell, S. Bron, S. Brouillet, C. V. Bruschi, B. Caldwell, V. Capuano, N. M. Carter, S. K. Choi, J. J. Codani, I. F. Connerton, A. Danchin, et al. 1997. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis.* Nature 390:249–256.
72. Kuzminov, A. 1995. Collapse and repair of replication forks in *Escherichia coli*. Mol. Microbiol. 16:373–384.
73. Loessner, M. J., R. B. Inman, P. Lauer, and R. Calendar. 2000. Complete nucleotide sequence, molecular analysis and genome structure of bacteriophage A118 of *Listeria monocytogenes*: implications for phage evolution. Mol. Microbiol. 35:324–340.
74. Low, K. B. and D. D. Porter. 1978. Modes of gene transfer and recombination in bacteria. Annu. Rev. Genet. 12:249–287.
75. Lucchini, S., F. Desiere, and H. Brüssow. 1998. The structural gene module in *Streptococcus thermophilus* bacteriophage ϕSfi11 shows a hierarchy of relatedness to *Siphoviridae* from a wide range of bacterial hosts. Virology 246:63–73.
76. Lurz, R., E. Orlova, D. Günther, P. Dube, A. Dröge, F. Weise, T. A. Trautner, M. van Heel, and P. Tavares. 2001. Structural organisation of the head-to-tail interface of bacteriophage SPP1. J. Mol. Biol. 310:1027–1037.
77. Maratea, D., R. M. Zsigray, and D. L. Balkwill. 1985. Characterization of *Bacillus subtilis* phage 41c. Curr. Microbiol. 12:261–266.
78. Marians, K. J. 2000. Replication and recombination intersect. Curr. Opin. Genet. Dev. 10:151–156.
79. Martínez-Jiménez, M., P. Mesa, and J. C. Alonso. 2002. *Bacillus subtilis* τ subunit of DNA polymerase III interacts with bacteriophage SPP1 hexameric G*40*P DNA helicase. Nucleic Acids Res. 30: 5056–5064.
80. Mastromei, G., S. Riva, A. Fietta, and L. Pagani. 1978. SPP1 DNA replicative forms: growth of phage SPP1 in *Bacillus*

subtilis mutants temperature-sensitive in DNA synthesis. Mol. Gen. Genet. 167:157–164.

81. McIntosh, P. K., R. Dunker, C. Mulder, and N. Brown. 1978. DNA of *Bacillus subtilis* bacteriophage SPP1: physical mapping and localization of the origin of replication. J. Virol. 28:865–876.
82. Mejean, V., and J. Claverys. 1988. Polarity of DNA entry in transformation of *Streptococcus pneumoniae*. Mol. Gen. Genet. 213: 444–448.
83. Mertens, G., E. Amann, and J. N. Reeve. 1979. Bacteriophage polypeptides synthesized in infected minicells and in vitro. Mol. Gen. Genet. 172:271–279.
84. Milanesi, G., and F. Melgara. 1974. In vivo and in vitro transcription of SPP1 DNA by host RNA polymerase. J. Virol. 14:1613–1614.
85. Missich, R., F. Weise, S. Chai, X. Pedré, R. Lurz, and J. C. Alonso. 1997. The replisome organizer (G*38*P) of *Bacillus subtilis* bacteriophage SPP1 forms specialized nucleoprotein complexes with two discrete distant regions of the SPP1 genome. J. Mol. Biol. 270:50–64.
86. Montenegro, M. A. and T. A. Trautner. 1981. In vivo transcription of *Bacillus subtilis* bacteriophage SPP1. Mol. Gen. Genet. 181:512–517.
87. Montenegro, M. A., H. Esche, and T. A. Trautner. 1976. Induction of mutations in *B. subtilis* phage SPP1 by growth on host cells carrying a mutator DNA polymerase III. Mol. Gen. Genet. 149:131–134.
88. Morelli, G., M. A. Montenegro, G. Hillenbrandt, E. Scherzinger, and T. A. Trautner. 1978. The genome of *B. subtilis* phage SPP1: assignment of 5′–3′-polarity to the complementary strands of SPP1 DNA. Mol. Gen. Genet. 164:93–97.
89. Morelli, G., C. Fisseau, B. Behrens, T. A. Trautner, J. Luh, S. W. Ratcliff, D. P. Allison, and A. T. Ganesan. 1979. The genome of *Bacillus subtilis* phage SPP1: the topology of DNA molecules. Mol. Gen. Genet. 168:153–161.
90. Murialdo, H. 1991. Bacteriophage λ DNA maturation and DNA packaging. Annu. Rev. Biochem. 60:125–153.
91. Newcomb, W. W., R. M. Juhas, D. R. Thomsen, F. L. Homa, A. D. Burch, S. K. Weller, and J. C. Brown. 2001. The UL6 gene product forms the portal for entry of DNA into the herpes simplex virus capsid. J. Virol. 75:10923–10932.
92. Novick, R.P., I. Edelman, and S. Lofdahl. 1986. Small *Staphylococcus aureus* plasmids are transduced as linear multimers that are formed and resolved by replicative processes. J. Mol. Biol. 192:209.
93. Orlova, E., P. Dube, E. Beckmann, F. Zemlin, R. Lurz, T. A. Trautner, P. Tavares, and M. van Heel. 1999. Structure of the 13-fold symmetric portal protein of bacteriophage SPP1. Nature Struct. Biol. 6:842–846.
94. Orlova, E. V., B. Gowen, A. Dröge, A. Stiege, R. Lurz, F. Weise, M. van Heel, and P. Tavares. 2003. Structure of a viral DNA gatekeeper at 10 Å resolution by cryo-electron microscopy. EMBO J. 22:1255–1262.
95. Pedré, X., F. Weise, S. Chai, G. Lüder, and J. C. Alonso. 1994. Analysis of *cis* and *trans* acting elements required for the initiation of DNA replication in the *Bacillus subtilis* bacteriophage SPP1. J. Mol. Biol. 236:1324–1340.
96. Piechowska, M., and M. S. Fox. 1971. Fate of transforming deoxyribonucleate in *Bacillus subtilis*. J. Bacteriol. 108:680–689.
97. Ratcliff, S. W., J. Luh, A. T. Ganesan, B. Behrens, R. Thompson, M. A. Montenegro, G. Morelli, and T. A. Trautner. 1979. The genome of *Bacillus subtilis* phage SPP1: the arrangement of restriction endonuclease generated fragments. Mol. Gen. Genet. 168:165.
98. Riva, S. 1969. Asymmetric transcription of *B. subtilis* phage SPP1 DNA in vitro. *Biochem. Biophys*. Res. Commun. 34: 824–830.
99. Riva, S., M. Polsinelli, and A. Falaschi. 1968. A new phage of *Bacillus subtilis* with infectious DNA having separable strands. J. Mol. Biol. 35:347–356.
100. Rottländer, E., and T. A. Trautner. 1970. Genetic and transfection studies with *B. subtilis* phage SP50. I. Phage mutants with restricted growth on *B. subtilis* strain 168. Mol. Gen. Genet. 108:47–60.
101. Santos, M. A. 1991. Bacteriófagos de *Bacillus subtilis* do grupo SPP1: características gerais, especificidade de adsorção e organização genómica. PhD thesis, Faculdade de Ciências da Universidade de Lisboa, Lisboa, Portugal.
102. Santos, M. A., H. de Lencastre, and L. J. Archer. 1983. *Bacillus subtilis* mutation blocking irreversible binding of bacteriophage SPP1. J. Gen. Microbiol. 129:3499–3504.
103. Santos, M. A., H. de Lencastre, and L. J. Archer. 1984. Homology between phages SPP1, 41c, 22a, ρ15 and SF6 of *Bacillus subtilis*. J. Gen. Virol. 65: 2067–2072.
104. Santos, M.A., J. Almeida, H. de Lencastre, G. Morelli, M. Kamke, and T. A. Trautner. 1986. Genomic organization of the related *Bacillus subtilis* bacteriophages SPP1, 41c, ρ15 and SF6. J. Virol. 60:702–707.
105. Schmidt, C., and H. Schmieger. 1984. Selective transduction of recombinant plasmids with cloned *pac* sites by *Salmonella* phage P22. Mol. Gen. Genet. 196:123–128.
106. Simpson, A. A., Y. Tao, P. G. Leiman, M. O. Badasso, Y. He, P. J. Jardine, N. H. Olson, M. C. Morais, S. Grimes, D. L. Anderson, T. S. Baker, and M. G. Rossmann. 2000. Structure of the bacteriophage ϕ29 DNA packaging motor. Nature 408:745–750.
107. Skalka, A. M. 1977. DNA replication: bacteriophage lambda. Curr. Top. Microbiol. Immunol. 78:201–237.
108. Soultanas, P., and D. B. Wigley. 2001. Unwinding the "Gordian knot" of helicase action. Trends Biochem. Sci. 26:47–54.
109. Spatz, H. C., and T. A. Trautner. 1970. One way to do experiments on gene conversion? Transfection with heteroduplex SPP1 DNA. Mol. Gen. Genet. 109:84–106.
110. Spatz, H. C., and T. A. Trautner. 1971. The role of recombination in transfection of *B. subtilis*. Mol. Gen. Genet. 113:174–190.
111. Stahl, F., and N. Murray. 1966. The evolution of gene clusters and genetic circularity in microorganisms. *Genetics* 53:569–576.
112. Stahl, F. W., I. Kobayashi, and M. M. Stahl. 1985. In phage λ, *cos* is a recombinator in the Red pathway. J. Mol. Biol. 181:199–209.
113. Stanley, E., G. F. Fitzgerald, C. Le Marrec, B. Fayard, and D. van Sinderen. 1997. Sequence analysis and

characterization of φO1205, a temperate bacteriophage infecting *Streptococcus thermophilus* CNRZ1205. Microbiology 143:3417–3429.

114. Steensma, H. Y., and J. Blok. 1979. Effect of calcium ions on the infection of *Bacillus subtilis* by bacteriophage SF6. J. Gen. Virol. 42:305–314.
115. Stiege, A., A. Isidro, A. Dröge, and P. Tavares. 2003. Specific and stoichiometric targetting of a DNA-binding protein to the SPP1 procapsid by interaction with the portal oligomer. Mol. Microbiol. 49:1201–1212.
116. Stüber, D., G. Morelli, H. Bujard, M. A. Montenegro, and T. A. Trautner. 1981. Promoter sites in the genome of *B. subtilis* phage SPP1. Mol. Gen. Genet. 181:518–521.
117. Tailor, R., G. Bensi, G. Morelli, U. Canosi, and T. A. Trautner. 1985. The genome of *Bacillus subtilis* phage SPP1: structure of an early promoter. J. Gen. Microbiol. 131:1259–1262.
118. Tavares, P., A. Dröge, R. Lurz, I. Graeber, E. Orlova, P. Dube, and M. van Heel. 1995. The SPP1 connection. FEMS Microbiol. Rev. 17:47–56.
119. Tavares, P., R. Lurz, A. Stiege, B. Rückert, and T. A. Trautner. 1996. Sequential headful packaging and fate of the cleaved DNA ends in bacteriophage SPP1. J. Mol. Biol. 264: 954–967.
120. Tavares, P., M. A. Santos, R. Lurz, G. Morelli, H. L. Lencastre, and T. A. Trautner. 1992. Identification of a gene in *Bacillus subtilis* bacteriophage SPP1 determining the amount of packaged DNA. J. Mol. Biol. 225:81–92.
121. Thuman-Commike, P. A., B. Greene, J. A. Malinski, J. King, and W. Chiu. 1998. Role of the scaffolding protein in P22 procapsid size determination suggested by $T = 4$ and $T = 7$ procapsid structures. Biophys. J. 74:559–568.
122. Trautner, T. A., and M. Noyer-Weidner. 1993. Restriction/modification and methylation systems, in *Bacillus subtilis*, related species and their phages, pp. 39–108. *In* A. Sonenshein, J. Hoch, and R. Losick (ed.) *Bacillus subtilis* and Other Gram-Positive Bacteria. American Society for Microbiology, Washington, D.C.
123. Trautner, T. A., and C. Spatz. 1973. Transfection in *B. subtilis*. Curr. Top. Microbiol. Immunol. 62:61–88.
124. Trautner, T. A., B. Pawlek, S. Bron, and C. Anagnostopoulos. 1974. Restriction and modification in *B. subtilis*. Biological aspects. Mol. Gen. Genet. 131:181–191.
125. Trautner, T. A., C. Spatz, B. Behrens, B. Pawlek, and M. Behnke. 1972. Exchange between complementary strands of DNA? Adv. Biosci. 8:79–87.
126. Valpuesta, J. M., and J. L. Carrascosa. 1994. Structure of viral connectors and their function in bacteriophage assembly and DNA packaging. Q. Rev. Biophys. 27:107–155.
127. van Heel, M., E. Orlova, P. Dube, and P. Tavares. 1996. Intrinsic versus imposed curvature in cyclical oligomers: the portal protein of bacteriophage SPP1. EMBO J. 15:4785–4688.
128. van Sinderen, D., H. Karsens, J. Kok, P. Terpstra, M. H. J. Ruiters, G. Venema, and A. Nauta. 1995. Sequence analysis and molecular characterization of the temperate lactococcal bacteriophage r1t. Mol. Microbiol. 19:1343–1355.
129. Viguera, E., P. Hernández, D. B. Krimer, A. S. Boitsov, R. Lurz, J. C. Alonso, and J. B. Schwartzman. 1996. The ColE1 unidirectional origin is a polar replication fork barrier. J. Biol. Chem. 271:22414–22421.
130. Viret, J. F., and J. C. Alonso. 1987. Generation of linear multigenome-length plasmid molecules in *Bacillus subtilis*. Nucleic Acids Res. 15:6349–6367.
131. Viret, J.F., A. Bravo, and J. C. Alonso. 1991. Recombination-dependent concatemeric plasmid replication. Microbiol. Rev. 55:675–683.
132. Weise, F., S. Chai, G. Lüder, and J. C. Alonso. 1994. Nucleotide sequence and complementation studies of the gene *35* region of the *Bacillus subtilis* bacteriophage SPP1. Virology 202:1046–1049.
133. Yamaguchi, T., T. Hayashi, H. Takami, K. Nakasone, M. Ohnishi, K. Nakayama, S. Yamada, H. Komatsuzawa, and M. Sugai. 2000. Phage conversion of exfoliative toxin A production in *Staphylococcus aureus*. Mol. Microbiol. 38:694–705.
134. Yasbin, R. E. and F. E. Young. 1974. Transduction in *Bacillus subtilis* by bacteriophage SPP1. J. Virol. 14:1343–1348.
135. Yasbin, R. E., V. C. Maino, and F. E. Young. 1976. Bacteriophage resistance in *Bacillus subtilis* 168, W23, and interstrain transformants. J. Bacteriol. 125:1120–1126.
136. Zsigray, R. M., A. L. Miss, and O. E. Landman. 1973. Penetration of a bacteriophage into *Bacillus subtilis*: blockage of infection by deoxyribonuclease. J. Virol. 11:69–77.

24

Bacteriophage P1

HANSJÖRG LEHNHERR

In the first edition of this volume, Michael B. Yarmolinsky and the late Nat L. Sternberg wrote a 147-page chapter filled to the brim with scientific, historical, and anecdotal information about bacteriophage P1 (198). Their work was the first comprehensive review covering 35 years of P1 research. Following in their footsteps, this review inevitably relies heavily on its predecessor. Any reader with an interest in P1 is strongly advised to also read the Yarmolinsky and Sternberg review, which extensively covers aspects of P1 biology that are still valid and up to date and therefore not covered here.

It is the aim of this account to provide the reader with a general overview of bacteriophage P1 biology, with special focus on developments during the last 15 years. The data are arranged to reflect a lytic phage infection cycle, with an inbuilt detour covering the plasmid life style of P1 during lysogeny.

Brief History

In 1951 Giuseppe Bertani managed to isolate three temperate bacteriophages from the *Escherichia coli* strain of Lisbonne and Carrère (14). The phages he baptized P1, P2, and P3 were serologically distinct and differed also in plaque size, with P1 showing the smallest plaques on a slightly pathogenic *Shigella dysenteriae* host. As plaque size was a very important feature for early studies on bacteriophages, Bertani himself opted against P1 and devoted much of his long life in science to the investigation of bacteriophage P2 (15). Phage P3 received very little attention and is nowadays almost completely forgotten. Bacteriophage P1 might have suffered a similar fate if not for Lennox (122), who found in 1955 that P1 is able to mediate generalized transduction. Since then P1 has been widely used as a tool to construct new bacterial strains with specific genotypes. Generalized transduction also provided an invaluable tool in the fine mapping of the *E. coli* chromosome in times when no genome sequencing efforts were possible. Accordingly, a stock of bacteriophage P1 can be found in almost every molecular biology laboratory around the world. Maybe as a consequence of this omnipresence, bacteriophage P1 served as a model organism in the study of many fundamental aspects of bacterial and phage biology, such as DNA restriction modification (8, 45), site-specific recombination (68), plasmid replication (24), partition (9), incompatibility (173), and addiction (118). Next to the bacteriophages T4 (chapter 18) and λ (chapter 27), P1 is among the best characterized today (119). Many aspects of its life cycle are understood in molecular details and also the nucleotide sequence of the entire P1 genome has recently been determined (124).

Early Infection

Morphology

Figure 24-1 shows an electron micrograph of a bacteriophage P1 particle. The phage genome is packaged into a head structure of icosahedral symmetry. The head is connected to an inflexible tail, formed by a rigid tail tube that is covered by a contractile sheath. The distal end of the tail is formed by a baseplate, to which six kinked tail fibers are attached (186). The general morphology of a P1 particle (187) is very similar to the morphology of other tailed bacteriophages such as Mu (chapter 30) or the group of T-even phages. Surprisingly though, there is very little amino acid sequence conservation found between these phages (106). The structural proteins of bacteriophage P1, with the exception of the tail fibers (64, 153), have not been characterized in detail.

Host Specificity

Bacteriophage P1 has a relatively wide host range, including a variety of Gram-negative species (198). The host specificity is governed by the *cix-cin* site-specific DNA inversion system (86, 99). As illustrated in figure 24-2, inversion of a 4.2 kb C-segment leads to the expression of alternate sets of tail

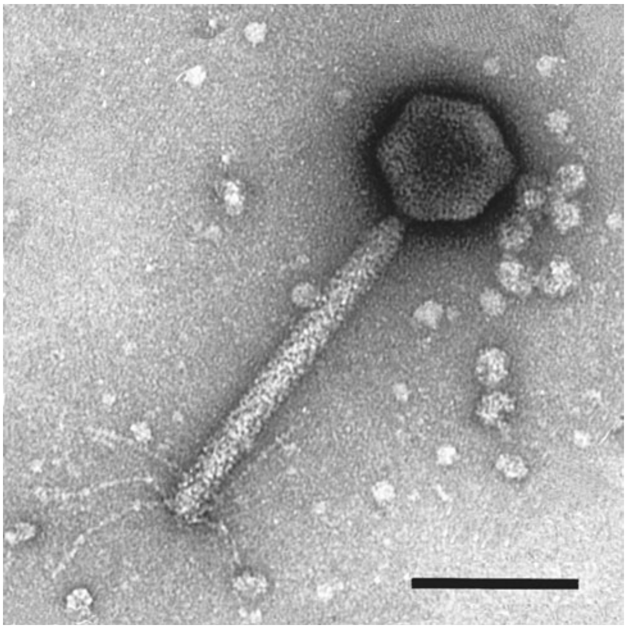


Figure 24-1 Electron micrograph of a bacteriophage P1 particle. The black bar represents 100 nm. Courtesy of M. Wurtz, Biocentre of the University of Basel, Switzerland.

fiber genes (R, S, U or R, S′, U′). Phages carrying RSU fibers were shown to be able to infect *E. coli* K12, *E. coli* B, and *E. coli* C, while phages carrying RS′U′ fibers do not normally infect these strains (97, 198). Inversion rarely occurs during lytic growth, but happens sufficiently frequently during lysogenic growth to completely randomize the orientation of the C-segment in a population. Thus phage stocks induced from a P1 lysogenic strain will contain equal numbers of particles with either host specificity (97, 99).

The precise inversion reaction is mediated by the 21.2 kDa Cin recombinase (67, 86). Crossing over occurs between two 26 bp *cix* sequences, flanking the C-segment. The *cix* sites contain imperfect inverted repeat sequences, which serve as recognition sites for the invertase protein (94, 95). Beside the Cin recombinase and the *cix* sites, also the host factor for inversion stimulation, FIS (67, 108, 110), and a 72 bp enhancer sequence, *sis*, located within the *cin* gene (90, 93), are required for efficient inversion. FIS binds to and bends the enhancer sequence (92, 93), providing the mould for the assembly of a precisely ordered synaptic Cin–*cix* complex, stimulating the efficiency of the inversion reaction while simultaneously preventing the formation of deletions (67).

Tail fiber variation by DNA inversion is not a unique feature of bacteriophage P1. Related systems include *gin* of phage Mu (105), *pin* of the defective prophage e14 (144), *min* of the defective plasmid-prophage p15B (100, 153, 154) and *ein* of Carotovoricin Er (140). In the bacterium *Salmonella typhimurium* an analogous *hin* system controls the expression of surface antigens (163). The respective recombinases show high degrees of similarity and, where tested, are functionally interchangeable (60).

Adsorption and Injection

The tail fibers promote the recognition of a potential host bacterium. The P1 receptor has been identified to be a terminal glucose moiety of the lipopolysaccharide core of the bacterial outer membrane (155). The interaction of at least three of the six tail fibers with specific receptor molecules is assumed to be sufficient to trigger the injection mechanism (34). The phage tail contracts, the tail tube is pushed through the baseplate, punctures the outer cell membrane and, according to most models (43), is also pushed through the bacterial cell wall. The latter process is facilitated by a phage-specific enzyme with mureolytic activity (116; H. Lehnherr, unpublished data), a so-called lytic transglycosylase (49). Previous studies linked the cell-wall-degrading activity of lytic transglycosylases to host cell lysis (49, 109), but recent evidence for the phages T7 and PRD1 demonstrated their role at the onset of an infection cycle (135, 151).

The content of the phage head is then injected into the periplasmic space of the host cell. The uptake of the P1 genome from the periplasm into the cytoplasm is mediated by an as yet uncharacterized pore in the inner membrane. Lobocka et al. (124) proposed that in analogy to the mechanism found in bacteriophage T7 (57, 176), a large internal head protein of bacteriophage P1, DarB (101), might form the required inner-membrane pore.

Another mechanism connected to the injection process is the *sim* (107) superinfection exclusion system. The Sim protein is synthesized as a 29.5 kDa precursor protein, and SecA-dependent cleavage of a 20 amino acid, hydrophobic

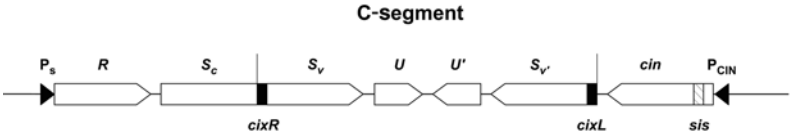


Figure 24-2 Organization of the DNA inversion system of P1. The two *cix* recombination sites flanking the invertible C-segment, and the enhancer sequence *sis*, located within the *cin* gene, are shown as black boxes. Two promoters, P_s, expressing the tail fiber operon and P_{CIN}, expressing the *cin* gene are shown as filled triangles. Genes are shown as open boxes with arrowheads indicating the direction of transcription. In the depicted orientation the three tail fiber proteins R, S, and U, are produced. A single inversion event would fuse Sc to Sv′, resulting in the expression of the three proteins R, S′, and U′.

leader peptide is essential for its function (129). Sim is rapidly expressed immediately after infection, and the mature 27.4 kDa protein is located in the inner membrane or the periplasmic space of the infected host cell (129). A plausible, though yet unproven hypothesis is that the Sim protein traps superinfecting P1 genomes in the periplasm, thus preventing them from interfering with the timing of an ongoing P1 infection cycle (107, 129).

Upon entering the cytoplasm the linear P1 DNA is rapidly circularized in order to avoid degradation of the phage DNA by cellular nucleases (174). Circularization is mediated either by Cre-dependent site-specific recombination using the presence of two *lox*-Cre sites in some phage genomes (87), by homologous recombination between the terminal redundant ends present in all P1 genomes or by other, yet uncharacterized mechanisms (198).

Restriction and Anti-restriction

Bacteriophage P1 specifies a type III restriction-modification system, which provides a P1 lysogen with a defence mechanism against superinfection by heterologous bacteriophages (17). Though the first of the three types of restriction-modification systems (17) to be discovered (8, 45), type III systems trailed in being characterized. It is mainly research in the last decade that has shed light on their activity. Two large proteins—the 74 kDa modification subunit, Mod, and the 111 kDa restriction subunit, Res (91)—form the *Eco*P1I restriction-modification enzyme complex, with a subunit composition of Res(2)Mod(2) (104). Sequence-specific binding to an asymmetric recognition sequence, AGÅCC (11), which is required for both methylation and restriction, is mediated by the modification subunit (6, 7, 66, 89). As a consequence of the asymmetry, only the adenine marked Å above is methylated by the Mod subunit, and the second strand of the recognition sequence, lacking an adenine, remains unmethylated. Figure 24-3 illustrates that upon replication of such hemimethylated sites, totally unmodified sites in the same orientation arise. For most restriction-modification systems, unmodified sites represent the signal to cleave DNA molecules that contain them (17). However, newly replicated DNA, of either P1 or host origin, is not a substrate for *Eco*P1I cleavage. This conceptual paradox was resolved when it was shown that two unmethylated sites, located in inverse orientation, are required for restriction activity (132, 133). The *Eco*P1I enzyme complex does bind to unmethylated sites and starts tracking along the DNA, using ATPase and helicase activities (104, 152, 199), but only the collision between two convergently tracking enzyme complexes triggers the actual cleavage reaction (104, 133).

The introduction of an *Eco*P1I restriction-modification system into a new host cell upon P1 infection is potentially suicidal for the phage, as the unchecked expression of restriction activity could result in the destruction of the unmethylated host genome (146). Especially the establishment of lysogeny would be severely hampered by such an activity. However, it was observed that while the modification activity is expressed immediately after DNA injection, restriction activity could only be detected with a considerable delay, allowing the modification subunit, which also acts as a monofunctional modification methyltransferase (6, 7), to completely methylate the host DNA (146). Such a sequential expression was not intuitively expected from the organization of the *mod* and *res* genes in a single operon (91, 98). Several lines of evidence now indicate that controls at both the transcriptional and, predominantly, the translational level are responsible for the delayed expression of the restriction subunit, providing the modification subunit with a head start (146, 160).

Newly injected bacteriophage P1 DNA is protected against degradation by type I restriction modification systems specified by Enterobacteriaceae (17, 101, 139). Responsible for this protection are two defences against restriction proteins: DarA and DarB. The DarA protein is synthesized as a precursor, which runs with an apparent molecular weight of 77 kDa during SDS-PAGE (175), though its calculated molecular weight is only 69.5 kDa (96). DarA is then processed by a particle maturation protease (175), and the mature form, with an apparent molecular weight of 68 kDa, is packaged into phage particles (175). DarB is also an

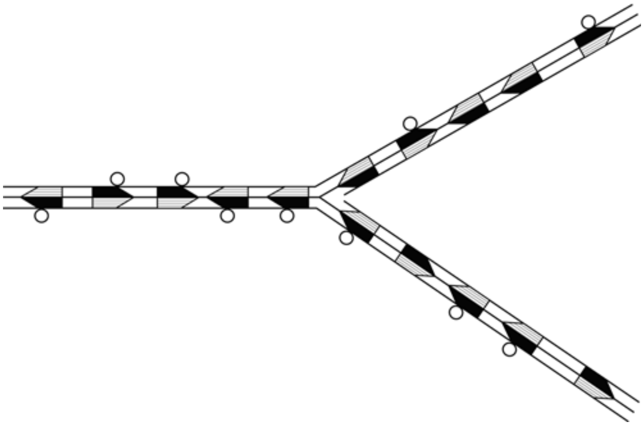


Figure 24-3 Replication fork in a P1 lysogen. The asymmetric *Eco*P1 recognition sequences are represented by differentially shaded arrows, with the arrowheads pointing in the direction of the cleavage sites. In the parental DNA molecule all *Eco*P1 sites are methylated (indicated by circles). In the two daughter molecules only those sites that inherited a methyl group are modified. However, as in each daughter molecule, all unmodified sites have the same orientation, no convergently tracking enzyme complexes will collide and thus the daughter molecules are not cleaved. *Eco*P1 enzyme complexes trailing the replication forks will then de novo modify the unmetylated sites, ensuring continued protection for the next round of replication (133). Modified from Meisel et al. (132).

internal head protein and both DarA and DarB are injected into the host cell alongside the phage DNA, where they act exclusively in *cis* (101). The Dar proteins do not directly inactivate the type I restriction-modification enzymes, and, as P1 DNA isolated from phage particles was not protected against type I restriction In vitro, Iida et al. (101) concluded that the protection was not due to a chemical modification either. Rather, as type I restriction enzymes track along the DNA from specific recognition sites to unspecific cleavage sites (17, 48), it was proposed that the Dar proteins act by sterically blocking these tracking movements. However, latest results indicated that DarB contains a domain which is highly similar to a DNA-methyltransferase (7, 124), and thus it can not be excluded that rapid methylation during the injection process, or immediately following it, might be at least partially responsible for the observed protection.

Lytic or Lysogenic Growth

Being a temperate bacteriophage, two life strategies are open to P1. It can either stably lysogenize the host and then replicate as a unit-copy number plasmid (102) or, alternatively, pursue a lytic growth cycle, resulting in the release of progeny phage particles (134). At the molecular level the decision between the two alternative pathways is regulated by the components of the complex tripartite immunity system, which has been thoroughly investigated by many members of the former research group of the late H. Schuster (80). Figure 24-4 provides a schematic illustration of the P1 immunity circuit with the three immunity regions *imm*C, *imm*I, and *imm*T. Immediately after injection of the P1 genome, seven P1 immunity functions are expressed. Their interplay, depending on the physiological conditions of the host, determines the outcome of the infection. Under optimal conditions with cells growing in rich medium at a low temperature of 20°C, up to 30% of the cells are lysogenized in *E. coli* (150).

*imm*C

The basic molecular switch is located in the *imm*C region of the P1 genome, containing two genes encoding the major repressor protein, C1 (44, 47), and the C1 repressor inactivator protein, Coi (83). The C1 repressor is a 32.5 kDa protein which binds to 22 operator sequences scattered widely over the P1 chromosome (81, 124). C1 recognizes asymmetric 17 bp sequences with the consensus ATTGCTCTAATAAATTT (81). The C1 repressor binds as a monomer to a single binding site and two monomers bind non-cooperatively to double binding sites (13, 82). Upon binding, C1 blocks the access of RNA polymerase to promoters located in the vicinity of the operator sites and thus represses the expression of genes associated with these promoters (81).

The Coi protein is a 7.7 kDa antirepressor protein which forms a 1:1 complex with the C1 repressor, thereby inactivating it (84). The C1 repressor has the potential to turn off the expression of Coi, as one of the C1 binding sites overlaps the promoter expressing the *coi*-*c*1 operon. The initial basic switch thus relies on the relative synthesis rate of the Coi and C1 proteins. An excess of Coi over C1 leads to lytic growth, while the opposite ratio promotes lysogeny.

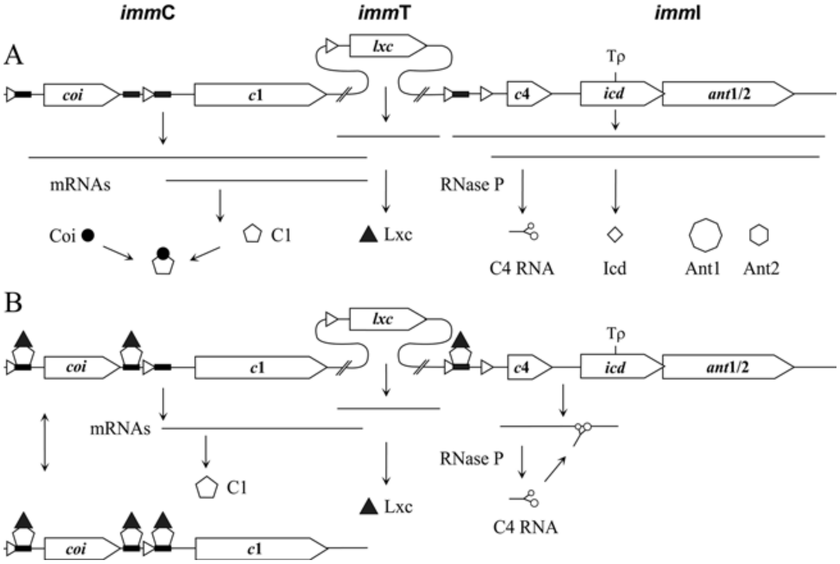


Figure 24-4 Organization of the tripartite P1 immunity system. The figure is not drawn to scale. Various promoter sequences, which are recognized by the *E. coli* RNA polymerase associated with σ^{70}, are indicated by triangles. Binding sites for the C1 repressor protein are shown by black bars.

*imm*I

Additional functions then further complicate the immunity switch. The antirepressor Ant, like Coi, can bind and inactivate C1, thus promoting lytic growth. Ant is a heterodimeric complex of two proteins, Ant1 (38.7 kDa) and Ant2 (29.0 kDa), which are encoded in the *imm*I region. Both subunits of Ant are expressed from a single gene, with the translation of the second protein starting at an in-frame start codon (147).

In order to establish lysogeny the expression of Ant has to be repressed and this goal is only partially achieved by the C1 repressor (see figure 24-4). Though there is a C1-controlled promoter reading toward the operon containing the genes *c*4, *icd* and *ant1/2*, there exists also a constitutive promoter expressing the same operon. Complete repression of Ant is achieved by the C4 RNA. This is one of the first antisense mechanisms ever described (25). The message of the *c*4 gene is processed by RNase P (71). The result is a 77 bp antisense RNA, which folds into a cloverleaf structure (26). This cloverleaf RNA then has the ability to bind to complementary sequences downstream in the same RNA message it has been processed from. This RNA–RNA interaction blocks the access of the ribosome to the Shine–Dalgarno (162) sequence of the *icd* gene. As the *ant1* gene is translationally coupled to *icd*, the same interaction also blocks Ant1 expression (26). However, the effect of C4 is not limited to translation, but also interferes with the transcription of the operon. The DNA-dependent RNA polymerase transcribing the operon pauses at a ρ-dependent terminator located within the *icd* gene (18). If C4 is present and prevents simultaneous translation of the growing message, then transcription is terminated. In the absence of C4 the translating ribosome alters the structure of the nascent mRNA chain and thereby triggers the RNA polymerase to complete transcription of the entire operon (18).

The *icd* gene not only acts as a mediator between *c*4 and *ant*, but also specifies a small 8.8 kDa toxic protein, which acts as a cell division inhibitor. Icd blocks division of the host cell until either lysogeny is successfully established or the entire developmental program of a lytic growth cycle is completed (148).

*imm*T

The corepressor protein, Lxc (formerly called Bof; 181), expressed from the *imm*T locus (157, 185), enhances the repression exerted by the C1 repressor protein and thus promotes lysogenic growth. Lxc is a small, 7.8 kDa protein which binds to DNA-bound C1 repressor and in such a DNA–C1–Lxc ternary complex increases the affinity of C1 for the operator sites (184). This activity of Lxc results in two antagonistic effects. As a consequence of the C1-autoregulatory control loop (figure 24-4 and discussed below), Lxc lowers the expression of *c*1, thus lowering the equilibrium concentration of the C1 protein within the cell. Lower C1 concentrations would result in reduced repression activity if this effect were not compensated for by the increased affinity of C1 for the operators. Whether the combined result of these two effects is a tighter repression or a slight derepression varies between different C1-regulated promoters, accounting for the very pleiotropic phenotype of *lxc* mutants (181, 185).

Maintenance of Lysogeny

In order to maintain lysogeny the intracellular concentration of the C1 repressor has to be buffered against fluctuations during growth. An efficient solution to this problem is an autoregulatory control loop as outlined in figure 24-4. One of the promoters expressing the *c*1 gene is controlled by the weakest C1-binding site found in the entire P1 genome (198). In the presence of Lxc, moderate or even low concentrations of C1 are sufficient to occupy this poor binding site, in addition to all other binding sites, thus switching off the expression of *c*1. Once the C1 concentration drops below a certain threshold level, this weak binding site will be the first to be clear of C1, allowing de novo synthesis in order to replenish the C1 pool before any other C1-controlled promoters get derepressed. The advantages of such a regulation are 2-fold. First, any unnecessary accumulation of C1 in the cell is avoided, reducing the metabolic burden of a P1 lysogen to its host. Second, the P1 prophage retains the ability to react quickly to stimuli that trigger lytic growth, as only small amounts of antirepressor protein would be sufficient to inactivate the low amounts of C1 present during equilibrium conditions.

Induction of Lytic Growth

The physiological conditions of the host cell are continuously monitored by the P1 prophage. Should changing circumstances disfavor the maintenance of lysogeny, a lytic growth cycle is initiated. Of the two antirepressor proteins Coi and Ant, only Ant can be expressed even in the presence of high concentrations of C1 and is thus able to overcome the C1 autoregulatory control loop. Conditions which change the overall RNase activity in the cell could impair the processing of the C4 RNA by RNase P (71), or could reduce the stability of the C4 RNA, thus leading to the expression of Ant. The inactivation of C1 by Ant then leads to the expression of all C1-regulated genes, including *coi*, triggering the switch from lysogenic to lytic growth. Such a model might reconcile the conflicting reports about the induction of P1 lysogens by UV light or other DNA-damaging agents (198). As UV damage does not directly affect the intracellular RNase activity levels, it is most likely that such a treatment only indirectly induces P1, possibly depending on variations in the bacterial host, as already suggested by Yarmolinsky and Sternberg (198).

Plasmid Maintenance

The P1 prophage does not integrate into the bacterial chromosome (102), but is maintained as a unit-copy plasmid within the cell. Mechanisms to increase plasmid stability, such as copy-number control, active partition, dimer resolution, or plasmid addiction, are important for any low-copy number plasmid, but P1 has the luxury of specifying all of them. Accordingly, the loss frequency of a P1 prophage is as low as 10^{-5} cured cells per generation (198).

Replication Control

In terms of understanding replication and its control, the P1 plasmid is one of the best studied members of the family of iteron-containing plasmids (24). Figure 24-5 shows the organization of the basic R replicon. The *repA* gene, encoding the 32.2 kDa phage-specific initiator protein, RepA, is flanked by the origin of plasmid replication, *oriR*, including the incompatibility control locus, *incC*, and a second incompatibility locus, *incA*. The hallmark of the replicon is the presence of 14 short repeated sequences, called iterons, forming both incompatibility loci. The iterons represent the binding sites of the RepA protein and show a 19 bp consensus sequence of GATGTGTGCTGGAGGGAAA (159). Binding of RepA to the iterons in *incC* is required for initiation, whereas binding to the *incA* iterons serves to control the plasmid copy number. In the absence of *incA*, the copy number increases about 10-fold, but replication is still controlled due to *incC* (143). Thus *incC* plays a dual role in both allowing replication and preventing overreplication.

Several host factors contribute to the initiation at *oriR*. Four GATC *dam* methylation sites within the origin itself need to be methylated (1, 2, 21). The methylation state of the origin was proposed to be negatively regulated by the host function, SeqA (22, 125), which binds to hemimethylated *dam* sites and sequesters them from the methylase (23). The melting of the DNA double strand at *oriR* occurs via the concerted action of RepA and the host factors, DnaA and HU (136, 190), thereby providing access for the host replication machinery (136, 142).

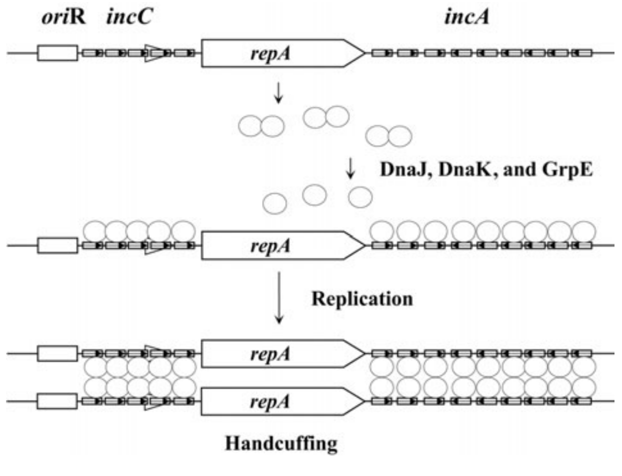


Figure 24-5 Organization of the P1 replication region. The figure is not drawn to scale. The 19 bp iteron sequences are indicated by small boxes. The arrows within the boxes indicate their directionality. An elongated triangle represents the promoter sequence expressing the *repA* gene.

The *repA* promoter is located within *incC* (figure 24-5) and is repressed upon RepA binding to the iterons (138). The expression of RepA is thus autoregulated. Freshly synthesized RepA protein apparently aggregates into dimers, but as only monomers show DNA-binding activity, there is a need for chaperones such as DnaK, DnaJ, and GrpE to convert the dimers into active monomers (35, 167, 168, 191). As RepA shows very little cooperativity while binding to *incC* (42), it is likely that the iterons located in *incA* will be at least partially occupied by RepA monomers before productive initiation occurs (figure 24-5). Thus *incA* could reduce copy number by simply titrating RepA. However, as an oversupply of RepA from artificial sources poorly overcomes the *incA* inhibition (24), a mechanism other than titration has to operate.

Such a mechanism is believed to be the pairing of origins (handcuffing) via bound RepA as illustrated in figure 24-5 (24, 137, 143). Handcuffed plasmid molecules are thought to be refractory to reinitiation due to steric hindrance. How handcuffing is reversed, other than by the separation of plasmid copies during partition, is not known, though an increase in RepA concentration has been shown to overcome a limited decrease in copy number, due to excess iterons (138). In summary, the copy number of P1 appears to be controlled by the availability of active RepA initiator protein, which promotes replication, but a continued increase in copy number is prevented by handcuffing (143).

Plasmid Partition

Replication control without active plasmid partition would contribute little to the stability of a low-copy-number plasmid. In P1 the replication and partition loci are physically linked (197). This circumstance, and the possibility that replication might be used to drive the partitioning of sibling plasmids, have suggested that the two processes are inseparable. However, this appears not to be the case, as evidence for active partitioning of unreplicated plasmids has been obtained (183). The P1 partition module is composed of two genes, *parA* and *parB*, and the centromere-like site *parS* (3). Our general understanding of the P1 partition process increased dramatically in the last decade due to major efforts in the laboratories led by Stuart J. Austin, Barbara E. Funnell, and Michael B. Yarmolinsky.

A minimal sequence of only 22 bp has been shown to have some partition activity (130), but a more complete *parS* sequence is approximately 94 bp in length and contains an asymmetrically located binding site for the integration host factor, IHF (182), which is flanked by highly conserved hexamer and heptamer boxes (38, 39, 52, 53). These boxes, with the consensus sequences TCGCCA and ATTTCA$^{C}/_{A}$, respectively, are the binding sites for the 37.4 kDa ParB protein (54, 123, 178). In a series of elegant experiments, some exploiting the subtle but distinct differences between the two partition systems of the closely related bacteriophages P1 and P7 (198), not only the species specificity but also the topology of the partition complex were worked out (53, 54, 74–77). As outlined in figure 24-6, ParB, in association with IHF, forms a high-affinity protein–DNA complex at *parS*, with the *parS* DNA wrapped around a core formed by the two proteins (20, 55, 179). Recent evidence demonstrated that ParB is able to pair partition sites (46), consistent with longstanding ideas about the partition mechanism (141).

The second partition protein, ParA (44.3 kDa), is an ATPase (36, 40, 41). In a complex with ADP, ParA exerts negative autoregulatory control over the *parAB* operon (36, 51, 77). Bound to ATP, ParA joins the partition complex at *parS*, via protein–protein interactions with ParB, and plays a direct and essential role in the partitioning process (19, 41). Both repressor and ATPase activities of ParA are stimulated in the presence of ParB (37).

The complex at *parS* also serves as a nucleation site for the polymerization of additional ParB molecules along the DNA flanking it on either side (149). Genes located in the vicinity of *parS* are transcriptionally silenced when they are covered by ParB (149, 195). However, how the silencing activity or any of the other interactions of the partition proteins at *parS* contribute to the actual equipartition of plasmid copies remains elusive. This holds true in spite of the latest localization studies, which revealed that P1 replication occurs in mid-cell and plasmid copies are then rapidly transported to the quarter and three quarter positions (61, 62). These positions match the intracellular localization of ParB (50) and, among other results, indicate that there are clear-cut differences between the partition systems of P1 and its host *E. coli* (56, 61).

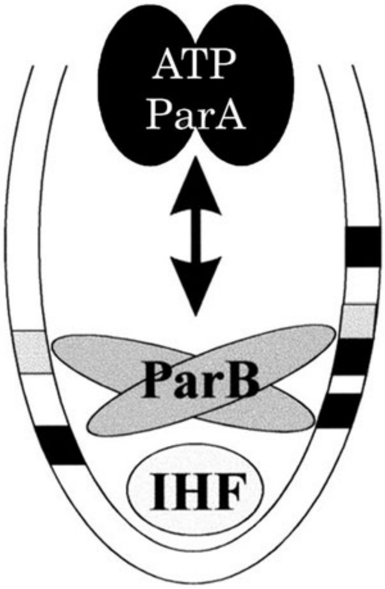


Figure 24-6 Partition complex at *parS*. The integration host factor, IHF (182), binds centrally within *parS* and introduces a strong bend. ParB is able to coordinately contact both conserved hexamer (gray boxes) and heptamer (black boxes) sequences, located in both arms of *parS*. ParA, complexed with ATP, then joins this core complex via interactions with ParB. For further details see text. Modified from Bouet et al. (20).

Dimer Resolution

Homologous recombination between sister plasmids can generate plasmid dimers, which would interfere with the partition machinery. Such dimers carry directly repeated *lox* sites and are efficiently resolved by the Cre recombinase (4). In vitro data implied that there might be an equilibrium between dimer formation and resolution mediated by Cre. However, in vivo results showed an almost exclusive preference for the dimer resolution reaction (5). In analogy to the *cer*-Xer dimer resolution system of ColE1 plasmids (177), the involvement of host factors, which favor the formation of a resolution complex (63, 88), might account for the observed differences between in vivo and In vitro results.

Plasmid Addiction

While the above three mechanisms attempt to avoid the loss of the P1 prophage, plasmid addiction is a way to retaliate should these measures fail. The P1 addiction module (118) contains two small genes, *phd* and *doc*, expressed from an autoregulated promoter (58, 127, 128). The 13.6 kDa Doc protein is a potent cell toxin, inhibiting the bacterial ribosome (78). In the presence of the P1 prophage the toxicity of Doc is neutralized by the 8.1 kDa antidote protein, Phd. Through differential expression of the *phd/doc* operon an excess of Phd over Doc is produced (118), which favors the formation of an inert heterotrimeric Phd(2)Doc complex (59). The antidote protein Phd is labile, as it is the substrate of the cellular ClpXP protease (121). Upon loss of the P1 prophage, ClpXP will degrade the residual antidote protein, leaving the stable toxin to kill plasmid-free cells (196).

Lytic Growth

Lytic Replication

The inactivation of the C1 repressor, either after infection or upon induction of a P1 lysogen, marks the onset of lytic growth. An entire set of genes, which is normally repressed by C1, is now expressed. Several of these so-called early functions play a role during lytic DNA replication and associated processes. Initiation of lytic replication occurs at *oriL*, mediated by the RepL initiator protein (30, 70, 170). Replication is bidirectional (29), initially of the theta type, but then predominantly of the sigma type during later stages of the infection cycle (28). The *repL* gene is coexpressed with *kilA*, coding for a cell division inhibitor similar to *icd* (148). Thus, bacterial cell division is blocked throughout lytic growth. The expression of the *kilA*/*repL* operon is controlled not only by C1 but also by an antisense regulatory mechanism to avoid inadvertent expression during lysogenic growth (79). The progress of the lytic replication fork is independent of the host DnaB protein (73), as P1 specifies a DnaB analog called Ban (85, 180). Other associated proteins, encoded by P1, include a Dam methylase (27, 33); a function, Ref (111, 126, 193, 194), that stimulates homologous recombination; a homolog of the DNA polymerase subunit, τ (124): HumD, a functional homolog of the *E. coli* UmuD′ protein involved in SOS mutagenesis (131); and a single-stranded DNA-binding protein (12, 113).

Activation of Late Transcription

During lytic growth only two major regulatory steps, early and late transcription, have been described for bacteriophage P1 (69, 120). Late genes, which encode the building blocks for the phage particle, as well as proteins involved in particle maturation, lysis control, and lysis (65, 96, 117, 158), are not expressed from standard *E. coli* promoters. Rather they are controlled by phage-specific late promoter sequences (65, 115), which share the −10 TATAAT consensus with *E. coli* promoters (72) but lack homology to the −35 TTGACA box (65). Conserved among the P1 late promoters is a sequence called −22 box, with the consensus ACAAGTTACTT, located 4 bp upstream of the −10 box (115). Such promoter sequences are not recognized per se by the *E. coli* σ^{70}-RNA polymerase, but need to be activated during lytic growth. A single phage-specific protein, Lpa (formerly called gp10), was identified to be essential for this activation (114). The expression of Lpa is controlled by C1 (120), defining it as an early function. Latest footprinting experiments showed that Lpa directly binds the −22 box in order to activate late transcription (69). Further experiments showed also that the RNA polymerase-associated host factor, SspA, is required for P1 late transcription (69, 103, 192). Exactly how Lpa and SspA cooperate in redirecting the host RNA polymerase towards P1 late promoter sequences is not clear.

Particle Morphogenesis

While most morphogenetic functions have been identified genetically, very little is known about the general assembly process of the P1 particle. Two internal head proteins, DarA and DarB, play a role in head morphogenesis (101), in addition to their involvement in defense against type I restriction described above. A protease, which is required for the proper maturation of phage heads (175), and a particle maturation function, *mat* (formerly called gene *1*), which shows a general defect in the production of infective particles (188), have been described (187). Unexpectedly, it was found that the *mat* gene is expressed both early and late during lytic growth (117), but the exact role that the Mat protein plays during particle maturation remains unclear.

Packaging

The initial packaging reaction starts at a specific cleavage site, *pac* (171). The pacase enzyme, formed by a heterodimer of the two proteins PacA (45.2 kDa) and PacB (55.6 kDa) (164, 166), binds and cleaves fully methylated *pac* sites (172) in association with the host factors, IHF and HU (165). The pacase then remains associated with one end of the cleaved *pac* site, interacts with a prohead, and mediates the incorporation of the phage DNA into the head. Subsequent *pac* sites present in the concatemeric DNA substrates, products of the sigma-type replication forks during the later stages of a lytic growth cycle (28), are not cleaved by the pacase. Such a processive "headful" packaging mechanism results in P1 heads being filled with DNA molecules larger than genome-size, with approximately 10% terminal redundancy (10, 198). The signal to ignore subsequent *pac* sites might be related to their methylation state (172) or, alternatively, it could be that the pacase is inhibited while it is associated with a prohead in the process of being filled. Evidence for the latter idea stems from the result that multiple rounds of *pac* cleavage are initiated when the packaged DNA molecule(s) does not have the required size to completely fill the phage head (31).

Transduction

The mechanism of formation of generalized transducing particles did not attract much attention, even though phage P1's ability to randomly transduce chromosomal markers is one of its most prominent features (198). P1 mutants with increased transduction frequencies have been isolated (96, 101, 189). Some contained mutations that were mapped to the *dar* operon and were thus linked to the head assembly process (96). Theoretically, mutations affecting either the availability of chromosomal-DNA free ends, the processivity of the packaging reaction, or the stability of transduced DNA molecules in the recipient cell all could result in an increased transduction frequency (198). Whether the respective protein(s) affects processivity via crosstalk between the head and the packaging machinery (165), or is associated with free ends of packaged DNA and thus stabilizes transduced DNA molecules, remains to be determined.

Cell Lysis

At the end of a lytic growth cycle newly formed progeny particles start to accumulate in the cytoplasm of the host. To release these progenies from their confinement, P1 specifies a 20.3 kDa, classical T4-type lysozyme (145, 200) with mureolytic activity (158). To secure the correct timing of cell lysis, the access of the lysozyme to the peptidoglycan layer is controlled by a holin, LydA (11.4 kDa), and a holin inhibitor, LydB (17.1 kDa) (96, 101, 158; chapter 10). The lysozyme cleaves the peptidoglycan strands, until the cell wall is no longer able to withstand the osmotic pressure within the cell. Lysis of the host cell terminates the lytic growth cycle and a burst of around 100–200 infective phage particles (134), like the one shown in figure 24-1, is released.

Outlook

The writing of this chapter marks the fiftieth anniversary of bacteriophage P1. Having reached such a venerable age, P1 research shows a few signs of slowing down, hand in hand with a general decrease in the interest in basic prokaryotic research. However, P1 genes continue to be prominent in the study of at least one major biological problem: the mechanism of partitioning (46). Also, many other aspects of P1 biology still pose unsolved puzzles, worthy of scientific scrutiny, and the recently completed nucleotide sequence of the P1 genome (124) might add a few more challenging problems to the list. On several occasions P1-derived systems also managed to transcend the limitations of prokaryotic research. A P1-based in vitro packaging system (169) is now widely used to clone large DNA fragments from almost any source (32, 161), and the *lox–cre* recombination system has turned out to be invaluable for manipulating the genomes of higher organisms (16, 156). Along these lines, P1 might well prove useful in yet other, maybe even medically relevant applications. A rational approach to phage therapy (chapter 48), for example, could be based on the ability of bacteriophage P1 to infect and destroy a broad range of pathogenic bacteria.

Acknowledgments

I would like to thank all members of the P1 community for the shared interest in and the endeavor to understand our fascinating model organism. I am particularly grateful to Thomas A. Bickle, Dhruba K. Chattoraj, Barbara E. Funnell, Tatiana V. Ilyina, Malgorzata B. Lobocka, Jürg Meyer, and Michael B. Yarmolinsky, for many fruitful discussions, constructive criticism, and encouragement that were vital to the completion of this manuscript.

References

1. Abeles, A. L., T. Brendler, and S. J. Austin. 1993. Evidence of two levels of control of P1 *oriR* and host *oriC* replication origins by DNA adenine methylation. J. Bacteriol. 175:7801–7807.
2. Abeles, A. L., and S. J. Austin. 1987. P1 plasmid replication requires methylated DNA. EMBO J. 6:3185–3189.
3. Abeles, A. L., S. A. Friedman, and S. J. Austin. 1985. Partition of unit-copy miniplasmids to daughter cells. III. The DNA sequence and functional organization of the P1 partition region. J. Mol. Biol. 185:261–272.
4. Abremski, K., R. Hoess, and N. Sternberg. 1983. Studies on the properties of P1 site-specific recombination: evidence for topologically unlinked products following recombination. Cell 32:1301–1311.
5. Adams, D. E., J. B. Bliska, and N. R. Cozzarelli. 1992. Cre-*lox* recombination in *Escherichia coli* cells. Mechanistic differences from the In vitro reaction. J. Mol. Biol. 226:661–673.
6. Ahmad, I., V. Krishnamurthy, and D. N. Rao. 1995. DNA recognition by the *Eco*P15I and *Eco*PI modification methyltransferases. Gene 157:143–147.
7. Ahmad, I., and D. N. Rao. 1996. Chemistry and biology of DNA methyltransferases. Crit. Rev. Biochem. Mol. Biol. 31:361–380.
8. Arber, W., and D. Dussoix. 1962. Host specificity of DNA produced by *Escherichia coli* I. Host controlled modification of bacteriophage λ J. Mol. Biol. 5:18–36.
9. Austin, S., and A. Abeles. 1983. Partition of unit-copy miniplasmids to daughter cells. I. P1 and F miniplasmids contain discrete, interchangeable sequences sufficient to promote equipartition. J. Mol. Biol. 169:353–372.
10. Bächi, B., and W. Arber. 1977. Physical mapping of *Bgl*II, *Bam*HI, *Eco*RI, *Hind*III and *Pst*I restriction fragments of bacteriophage P1 DNA. Mol. Gen. Genet. 153:311–324.
11. Bächi, B., J. Reiser, and V. Pirrotta. 1979. Methylation and cleavage sequences of the *Eco*P1 restriction-modification enzyme. J. Mol. Biol. 128:143–163.
12. Bendtsen, J. D., A. S. Nilsson, and H. Lehnherr. 2002. Phylogenetic and functional analysis of the bacteriophage P1 single-stranded DNA-binding protein. Unpublished results.
13. Bergmann, J., D. Rünzler, H. Lehnherr, G. Köhler, C. Kratky, and W. Keller. 2002. DNA-binding characteristics of the major repressor protein, C1, of bacteriophage P1. Unpublished results.
14. Bertani, L. E. 1951. Studies in lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. J. Bacteriol. 62:293–300.
15. Bertani, L. E., and G. Bertani. 1971. Genetics of P2 and related phages. Adv. Genet. 16:199–237.
16. Bethke, B. D., and B. Sauer. 2000. Rapid generation of isogenic mammalian cell lines expressing recombinant transgenes by use of Cre recombinase. Methods Mol. Biol. 133:75–84.
17. Bickle, T. A., and D. H. Krüger. 1993. Biology of DNA restriction. Microbiol. Rev. 57:434–450.
18. Biere, A. L., M. Citron, and H. Schuster. 1992. Transcriptional control via translational repression by *c*4 antisense RNA of bacteriophages P1 and P7. Genes Dev. 6:2409–2416.

19. Bouet, J. Y., and B. E. Funnell . 1999. P1 ParA interacts with the P1 partition complex at *parS* and an ATP–ADP switch controls ParA activities. EMBO J. 18:1415–1424.
20. Bouet, J. Y., J. A. Surtees, and B. E. Funnell. 2000. Stoichiometry of P1 plasmid partition complexes. J. Biol. Chem. 275:8213–8219.
21. Brendler, T., A. Abeles, and S. Austin. 1991. Critical sequences in the core of the P1 plasmid replication origin. J. Bacteriol. 173:3935–3942.
22. Brendler, T., A. Abeles, and S. Austin. 1995. A protein that binds to the P1 origin core and the *oriC* 13mer region in a methylation-specific fashion is the product of the host *seqA* gene. EMBO J. 14:4083–4089.
23. Brendler, T., and S. Austin. 1999. Binding of SeqA protein to DNA requires interaction between two or more complexes bound to separate hemimethylated GATC sequences. EMBO J. 18:2304–2310.
24. Chattoraj, D. K. 2000. Control of plasmid DNA replication by iterons: no longer paradoxical. Mol. Microbiol. 37:467–476.
25. Citron, M., and H. Schuster. 1990. The *c*4 repressors of bacteriophages P1 and P7 are antisense RNAs. Cell 62:591–598.
26. Citron, M., and H. Schuster. 1992. The *c*4 repressor of bacteriophage P1 is a processed 77 base antisense RNA. Nucleic Acids Res. 20:3085–3090.
27. Citron, M., M. Velleman, and H. Schuster. 1989. Three additional operators, Op21, Op68, and Op88, of bacteriophage P1. Evidence for control of the P1 *dam* methylase by Op68. J. Biol. Chem. 264:3611–3617.
28. Cohen, G. 1983. Electron microscopy study of early lytic replication forms of bacteriophage P1 DNA. Virology 131:159–170.
29. Cohen, G., E. Or, W. Minas, and N. L. Sternberg. 1996. The bacteriophage P1 lytic replicon: directionality of replication and *cis*-acting elements. Gene 175:151–155.
30. Cohen, G., and N. L. Sternberg. 1989. Genetic analysis of the lytic replication of bacteriophage P1. I. Isolation and partial characterization. J. Mol. Biol. 207:99–109.
31. Coren, J. S., J. C. Pierce, and N. Sternberg. 1995. Headful packaging revisited: the packaging of more than one DNA molecule into a bacteriophage P1 head. J. Mol. Biol. 249:176–184.
32. Coren, J. S., and N. Sternberg. 2001. Construction of a PAC vector system for the propagation of genomic DNA in bacterial and mammalian cells and subsequent generation of nested deletions in individual library members. Gene 264:11–18.
33. Coulby, J. N., and N. L. Sternberg. 1988. Characterization of the phage P1 *dam* gene. Gene 74:191.
34. Crawford, J. T., and E. B. Goldberg. 1980. The function of tail fibers in triggering baseplate expansion of bacteriophage T4. J. Mol. Biol. 139:679–690.
35. DasGupta, S., G. Mukhopadhyay, P. P. Papp, M. S. Lewis, and D. K. Chattoraj. 1993. Activation of DNA binding by the monomeric form of the P1 replication initiator RepA by heat shock proteins DnaJ and DnaK. J. Mol. Biol. 232:23–34.
36. Davey, M. J., and B. E. Funnell. 1994. The P1 plasmid partition protein ParA. A role for ATP in site-specific DNA binding. J. Biol. Chem. 269:29908–29913.
37. Davey, M. J., and B. E. Funnell. 1997. Modulation of the P1 plasmid partition protein ParA by ATP, ADP, and P1 ParB. J. Biol. Chem. 272:15286–15292.
38. Davis, M. A., and S. J. Austin. 1988. Recognition of the P1 plasmid centromere analog involves binding of the ParB protein and is modified by a specific host factor. EMBO J. 7:1881–1888.
39. Davis, M. A., K. A. Martin, and S. J. Austin. 1990. Specificity switching of the P1 plasmid centromere-like site. EMBO J. 9:991–998.
40. Davis, M. A., K. A. Martin, and S. J. Austin. 1992. Biochemical activities of the parA partition protein of the P1 plasmid. Mol. Microbiol. 6:1141–1147.
41. Davis, M. A., L. Radnedge, K. A. Martin, F. Hayes, B. Youngren, and S. J. Austin. 1996. The P1 ParA protein and its ATPase activity play a direct role in the segregation of plasmid copies to daughter cells. Mol. Microbiol. 21:1029–1036.
42. Dibbens, J. A., K. Muraiso, and D. K. Chattoraj. 1997. Chaperone-mediated reduction of RepA dimerization is associated with RepA conformational change. Mol. Microbiol. 26:185–195.
43. Dreiseikelmann, B. 1994. Translocation of DNA across bacterial membranes. Microbiol. Rev. 58:293–316.
44. Dreiseikelmann, B., M. Velleman, and H. Schuster. 1988. The *c*1 repressor of bacteriophage P1. J. Biol. Chem. 263:1391–1397.
45. Dussoix, D., and W. Arber. 1962. Host specificity of DNA produced by *Escherichia coli* I. Control of acceptance of DNA from infecting phage λ J. Mol. Biol. 5:37–49.
46. Edgar, R., D. K. Chattoraj, and M. B. Yarmolinsky. 2001. Pairing of P1 plasmid partition sites by ParB. Mol. Microbiol. 42:1363–1370.
47. Eliason, J. L., and N. L. Sternberg. 1987. Characterization of the binding sites of *c*1 repressor of bacteriophage P1. Evidence for multiple asymmetric sites. J. Mol. Biol. 198:281–293.
48. Endlich, B., and S. Linn. 1985. The DNA restriction endonuclease of *Escherichia coli* B. I. Studies of the DNA translocation and the ATPase activities. J. Biol. Chem. 260:5720–5728.
49. Engel, H., B. Kazemier, and W. Keck. 1991. Murein-metabolizing enzymes from *Escherichia coli*: sequence analysis and controlled overexpression of the *slt* gene, which encodes the soluble lytic transglycosylase. J. Bacteriol. 173:6773–6782.
50. Erdmann, N., T. Petroff, and B. E. Funnell. 1999. Intracellular localization of P1 ParB protein depends on ParA and *parS*. Proc. Natl. Acad. Sci. USA 96:14905–14910.
51. Fung, E., J. Y. Bouet, and B. E. Funnell. 2001. Probing the ATP-binding site of P1 ParA: partition and repression have different requirements for ATP binding and hydrolysis. EMBO J. 20:4901–4911.
52. Funnell, B. E. 1988. Participation of *Escherichia coli* integration host factor in the P1 plasmid partition system. Proc. Natl. Acad. Sci. USA 85:6657–6661.
53. Funnell, B. E. 1991. The P1 plasmid partition complex at *parS*. The influence of *Escherichia coli* integration host factor and of substrate topology. J. Biol. Chem. 266:14328–14337.

54. Funnell, B. E., and L. Gagnier. 1993. The P1 plasmid partition complex at *parS*. II. Analysis of ParB protein binding activity and specificity. J. Biol. Chem. 268: 3616–3624.
55. Funnell, B. E., and L. Gagnier. 1994. P1 plasmid partition: binding of P1 ParB protein and *Escherichia coli* integration host factor to altered *parS* sites. Biochimie 76:924–932.
56. Funnell, B. E., and L. Gagnier. 1995. Partition of P1 plasmids in *Escherichia coli mukB* chromosomal partition mutants. J. Bacteriol. 177:2381–2386.
57. Garcia, L. R., and I. J. Molineux. 1996. Transcription-independent DNA translocation of bacteriophage T7 DNA into *Escherichia coli*. J. Bacteriol. 178:6921–6929.
58. Gazit, E., and R. T. Sauer. 1999. Stability and DNA binding of the *phd* protein of the phage P1 plasmid addiction system. J. Biol. Chem. 274:2652–2657.
59. Gazit, E., and R. T. Sauer. 1999. The Doc toxin and Phd antidote proteins of the bacteriophage P1 plasmid addiction system form a heterotrimeric complex. J. Biol. Chem. 274:16813–16818.
60. Glasgow, A. C., K. T. Hughes, and M. I. Simon. 1989. Bacterial DNA inversion systems, pp. 637–659. *In* D. E. Berg and M. M. Howe (eds.) Mobile DNA. Amercian Society for Microbiology, Washington, D.C.
61. Gordon, G. S., D. Sitnikov, C. D. Webb, A. Teleman, A. Straight, R. Losick, A. W. Murray, and A. Wright. 1997. Chromosome and low copy plasmid segregation in *E. coli*: visual evidence for distinct mechanisms. Cell 90:1113–1121.
62. Gordon, G. S., and A. Wright. 2000. DNA segregation in bacteria. Annu. Rev. Microbiol. 54:681–708.
63. Guhathakurta, A., I. Viney, and D. Summers. 1996. Accessory proteins impose site selectivity during ColE1 dimer resolution. Mol. Microbiol. 20:613–620.
64. Guidolin, A., J. M. Zingg, and W. Arber. 1989. Organization of the bacteriophage P1 tail-fibre operon. Gene 76:239–243.
65. Guidolin, A., J. M. Zingg, H. Lehnherr, and W. Arber. 1989. Bacteriophage P1 tail-fibre and *dar* operons are expressed from homologous phage-specific late promoter sequences. J. Mol. Biol. 208:615–622.
66. Hadi, S. M., B. Bächi, S. Iida, and T. A. Bickle. 1983. DNA restriction-modification enzymes of phage P1 and plasmid p15B. Subunit functions and structural homologies. J. Mol. Biol. 165:19–34.
67. Haffter, P., and T. A. Bickle. 1987. Purification and DNA-binding properties of FIS and Cin, two proteins required for the bacteriophage P1 site-specific recombination system, *cin*. J. Mol. Biol. 198:579–587.
68. Hamilton, D. L., and K. Abremski. 1984. Site-specific recombination by the bacteriophage P1 *lox*–Cre system. Cre-mediated synapsis of two *lox* sites. J. Mol. Biol. 178:481–486.
69. Hansen, A.-M., H. Lehnherr, X. Wang, V. Mobley, and D. J. Jin. 2003. *Escherichia coli* SspA is a transcription activator for bacteriophage P1 late genes. Mol. Microbiol. 48: 1621–1631.
70. Hansen, E. B. 1989. Structure and regulation of the lytic replicon of phage P1. J. Mol. Biol. 207:135–149.
71. Hartmann, R. K., J. Heinrich, J. Schlegl, and H. Schuster. 1995. Precursor of C4 antisense RNA of bacteriophages P1 and P7 is a substrate for RNase P of *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A 92:5822–5826.
72. Hawley, D. K., and W. R. McClure. 1983. Compilation and analysis of *Escherichia coli* promoter and DNA sequences. Nucleic Acids Res. 11:2237–2255.
73. Hay, N., and G. Cohen. 1983. Requirement of *E. coli* DNA synthesis functions for the lytic replication of bacteriophage P1. Virology 131:193–206.
74. Hayes, F., and S. Austin. 1994. Topological scanning of the P1 plasmid partition site. J. Mol. Biol. 243:190–198.
75. Hayes, F., and S. J. Austin. 1993. Specificity determinants of the P1 and P7 plasmid centromere analogs. Proc. Natl. Acad. Sci. USA 90:9228–9232.
76. Hayes, F., M. A. Davis, and S. J. Austin. 1993. Fine-structure analysis of the P7 plasmid partition site. J. Bacteriol. 175:3443–3451.
77. Hayes, F., L. Radnedge, M. A. Davis, and S. J. Austin. 1994. The homologous operons for P1 and P7 plasmid partition are autoregulated from dissimilar operator sites. Mol. Microbiol. 11:249–260.
78. Hazan, R., B. Sat, M. Reches, and H. Engelberg-Kulka. 2001. Postsegregational killing mediated by the P1 phage "addiction module" *phd-doc* requires the *Escherichia coli* programmed cell death system *mazEF*. J. Bacteriol. 183:2046–2050.
79. Heinrich, J., H.-D. Riedel, B. Rückert, R. Lurz, and H. Schuster. 1995. The lytic replicon of bacteriophage P1 is controlled by an antisense RNA. Nucleic Acids Res. 23:1468–1474.
80. Heinrich, J., M. Velleman, and H. Schuster. 1995. The tripartite immunity system of phages P1 and P7. FEMS Microbiol. Rev. 17:121–126.
81. Heinzel, T., M. Velleman, and H. Schuster. 1989. *ban* operon of bacteriophage P1. Mutational analysis of the *c*1 repressor-controlled operator. J. Mol. Biol. 205:127–135.
82. Heinzel, T., M. Velleman, and H. Schuster. 1989. *ban* operon of bacteriophage P1. Mutational analysis of the *c*1 repressor-controlled operator. J. Mol. Biol. 205:127–135.
83. Heinzel, T., M. Velleman, and H. Schuster. 1990. The *c*1 repressor inactivator protein *coi* of bacteriophage P1. Cloning and expression of *coi* and its interference with *c*1 repressor function. J. Biol. Chem. 265:17928–17934.
84. Heinzel, T., M. Velleman, and H. Schuster. 1992. C1 repressor of phage P1 is inactivated by noncovalent binding of P1 Coi protein. J. Biol. Chem. 267:4183–4188.
85. Heisig, A., I. Severin, A. K. Seefluth, and H. Schuster. 1987. Regulation of the *ban* gene containing operon of prophage P1. Mol. Gen. Genet. 206:368–376.
86. Hiestand-Nauer, R., and S. Iida. 1983. Sequence of the site-specific recombinase gene cin and of its substrates serving in the inversion of the C segment of bacteriophage P1. EMBO J. 2:1733–1740.
87. Hochman, L., N. Segev, N. Sternberg, and G. Cohen. 1983. Site-specific recombinational circularization of bacteriophage P1 DNA. Virology 131:11–17.
88. Hodgman, T. C., H. Griffiths, and D. K. Summers. 1998. Nucleoprotein architecture and ColE1 dimer resolution: a hypothesis. Mol. Microbiol. 29:545–558.
89. Hornby, D. P., M. Müller, and T. A. Bickle. 1987. High level expression of the *Eco*P1 modification methylase gene and characterisation of the gene product. Gene 54:239–245.

90. Huber, H. E., S. Iida, W. Arber, and T. A. Bickle. 1985. Site-specific DNA inversion is enhanced by a DNA sequence element in *cis*. Proc. Natl. Acad. Sci. USA 82:3776–3780.
91. Humbelin, M., B. Suri, D. N. Rao, D. P. Hornby, H. Eberle, T. Pripfl, S. Kenel, and T. A. Bickle. 1988. Type III DNA restriction and modification systems *Eco*P1 and *Eco*P15. Nucleotide sequence of the *Eco*P1 operon, the *Eco*P15 *mod* gene and some *Eco*P1 *mod* mutants. J. Mol. Biol. 200:23–29.
92. Hübner, P., and W. Arber. 1989. Mutational analysis of a prokaryotic recombinational enhancer element with two functions. EMBO J. 8:577–585.
93. Hübner, P., P. Haffter, S. Iida, and W. Arber. 1989. Bent DNA is needed for recombinational enhancer activity in the site-specific recombination system Cin of bacteriophage P1. The role of FIS protein. J. Mol. Biol. 205:493–500.
94. Iida, S., and R. Hiestand-Nauer . 1986. Localized conversion at the crossover sequences in the site-specific DNA inversion system of bacteriophage P1. Cell 45:71–79.
95. Iida, S., and R. Hiestand-Nauer. 1987. Role of the central dinucleotide at the crossover sites for the selection of quasi sites in DNA inversion mediated by the site-specific Cin recombinase of phage P1. Mol. Gen. Genet. 208:464–468.
96. Iida, S., R. Hiestand-Nauer, H. Sandmeier, H. Lehnherr, and W. Arber. 1998. Accessory genes in the *darA* operon of bacteriophage P1 affect anti-restriction function, generalized transduction, head morphogenesis and host cell lysis. Virology 251:49–58.
97. Iida, S., H. Huber, R. Hiestand-Nauer, J. Meyer, T. A. Bickle, and W. Arber. 1984. The bacteriophage P1 site-specific recombinase Cin: recombination events and DNA recognition sequences. Cold Spring Harb. Symp. Quant. Biol. 49:769–777.
98. Iida, S., J. Meyer, B. Bächi, M. Stalhammar-Carlemalm, S. Schrickel, T. A. Bickle, and W. Arber. 1983. DNA restriction-modification genes of phage P1 and plasmid p15B. Structure and In vitro transcription. J. Mol. Biol. 165:1–18.
99. Iida, S., J. Meyer, K. E. Kennedy, and W. Arber. 1982. A site-specific, conservative recombination system carried by bacteriophage P1. Mapping the recombinase gene *cin* and the cross-over sites *cix* for the inversion of the C segment. EMBO J. 1:1445–1453.
100. Iida, S., H. Sandmeier, P. Hübner, R. Hiestand-Nauer, K. Schneitz, and W. Arber. 1990. The Min DNA inversion enzyme of plasmid p15B of *Escherichia coli* 15T$^-$: a new member of the Din family of site-specific recombinases. Mol. Microbiol. 4:991–997.
101. Iida, S., M. B. Streiff, T. A. Bickle, and W. Arber. 1987. Two DNA antirestriction systems of bacteriophage P1, *darA*, and *darB*: characterization of *darA*$^-$ phage. Virology 157:156–166.
102. Ikeda, H., and J. Tomizawa. 1968. Prophage P1, and extrachromosomal replication unit. Cold Spring Harb. Symp. Quant. Biol. 33:791–798.
103. Ishihama, A., and T. Saitoh. 1979. Subunits of RNA polymerase in function and structure. IX. Regulation of RNA polymerase activity by stringent starvation protein (SSP). J. Mol. Biol. 129:517–530.
104. Janscak, P., U. Sandmeier, M. D. Szczelkun, and T. A. Bickle. 2001. Subunit assembly and mode of DNA cleavage of the type III restriction endonucleases *Eco*P1I and *Eco*P15I. J. Mol. Biol. 306:417–431.
105. Kamp, D., R. Kahmann, D. Zipser, T. R. Broker, and L. T. Chow. 1978. Inversion of the G DNA segment of phage Mu controls phage infectivity. Nature 271:577–580.
106. Klaus, S., D. Krüger, and J. Meyer. 1992. Bakterienviren. Gustav Fischer Verlag, Jena.
107. Kliem, M., and B. Dreiseikelmann. 1989. The superimmunity gene *sim* of bacteriophage P1 causes superinfection exclusion. Virology 171:350–355.
108. Koch, C., and R. Kahmann. 1986. Purification and properties of the *Escherichia coli* host factor required for inversion of the G segment in bacteriophage Mu. J. Biol. Chem. 261:15673–15678.
109. Koonin, E. V., and K. E. Rudd. 1994. A conserved domain in putative bacterial and bacteriophage transglycosylases. Trends Biochem. Sci. 19:106–107.
110. Kostrewa, D., J. Granzin, C. Koch, H. W. Choe, S. Raghunathan, W. Wolf, J. Labahn, R. Kahmann, and W. Saenger . 1991. Three-dimensional structure of the *E. coli* DNA-binding protein FIS. Nature 349:178–180.
111. Laufer, C. S., J. B. Hays, B. E. Windle, T. S. Schaefer, E. H. Lee, S. L. Hays, and M. R. McClure. 1989. Enhancement of *Escherichia coli* plasmid and chromosomal recombination by the Ref function of bacteriophage P1. Genetics 123:465–476.
112. Lehnherr, H. 2002. unpublished data.
113. Lehnherr, H., J. D. Bendtsen, F. Preuss, and T. V. Ilyina. 1999. Identification and characterization of the single-stranded DNA-binding protein of bacteriophage P1. J. Bacteriol. 181:6463–6468.
114. Lehnherr, H., A. Guidolin, and W. Arber. 1991. Bacteriophage P1 gene *10* encodes a *trans*-activating factor required for late gene expression. J. Bacteriol. 173:6438–6445.
115. Lehnherr, H., A. Guidolin, and W. Arber. 1992. Mutational analysis of the bacteriophage P1 late promoter sequence Ps. J. Mol. Biol. 228:101–107.
116. Lehnherr, H., A.-M. Hansen, and T. V. Ilyina. 1998. Penetration of the bacterial cell wall: a family of lytic transglycosylases in bacteriophages and conjugative plasmids. Mol. Microbiol. 30:454–457.
117. Lehnherr, H., C. D. Jensen, A. R. Stenholm, and A. Dueholm. 2001. Dual regulatory control of a particle maturation function of bacteriophage P1. J. Bacteriol. 183:4105–4109.
118. Lehnherr, H., E. Maguin, S. Jafri, and M. B. Yarmolinsky. 1993. Plasmid addiction genes of bacteriophage P1: *doc*, which causes cell death on curing of prophage, and *phd*, which prevents host death when prophage is retained. J. Mol. Biol. 233:414–428.
119. Lehnherr, H., and J. Meyer. 1999. Enterobacteria phage P1 (*Myoviridae*), pp. 455–461. *In* A. Granoff and R. G. Webster (eds.) Encyclopedia of Virology. Academic Press, London.
120. Lehnherr, H., M. Velleman, A. Guidolin, and W. Arber. 1992. Bacteriophage P1 gene *10* is expressed from a promoter-operator sequence controlled by C1 and Bof proteins. J. Bacteriol. 174:6138–6144.

121. Lehnherr, H., and M. B. Yarmolinsky. 1995. Addiction protein Phd of plasmid prophage P1 is a substrate of the ClpXP serine protease of *Escherichia coli*. Proc. Natl. Acad. Sci. USA 92:3274–3277.
122. Lennox, E. S. 1955. Transduction of linked genetic characters of the host by bacteriophage P1. Virology 1:190–206.
123. Lobocka, M., and M. B. Yarmolinsky. 1996. P1 plasmid partition: a mutational analysis of ParB. J. Mol. Biol. 259:366–382.
124. Lobocka, M. B., D. Rose, M. Rusin, A. Samojedny, M. B. Yarmolinsky, H. Lehnherr, and F. C. Blattner. 2002. The genome of bacteriophage P1: analysis based on the complete nucleotide sequence of two related isolates. Unpublished results.
125. Lu, M., J. L. Campbell, E. Boye, and N. Kleckner. 1994. SeqA: a negative modulator of replication initiation in *E. coli*. Cell 77:413–426.
126. Lu, S. D., D. Lu, and M. Gottesman. 1989. Stimulation of IS*1* excision by bacteriophage P1 *ref* function. J. Bacteriol. 171:3427–3432.
127. Magnuson, R., H. Lehnherr, G. Mukhopadhyay, and M. B. Yarmolinsky. 1996. Autoregulation of the plasmid addiction operon of bacteriophage P1. J. Biol. Chem. 271:18705–18710.
128. Magnuson, R., and M. B. Yarmolinsky. 1998. Corepression of the P1 addiction operon by Phd and Doc. J. Bacteriol. 180:6342–6351.
129. Maillou, J., and B. Dreiseikelmann. 1990. The *sim* gene of *Escherichia coli* phage P1: nucleotide sequence and purification of the processed protein. Virology 175:500–507.
130. Martin, K. A., M. A. Davis, and S. Austin. 1991. Fine-structure analysis of the P1 plasmid partition site. J. Bacteriol. 173:3630–3634.
131. McLenigan, M. P., O. I. Kulaeva, D. G. Ennis, A. S. Levine, and R. Woodgate. 1999. The bacteriophage P1 HumD protein is a functional homolog of the prokaryotic UmuD′-like proteins and facilitates SOS mutagenesis in *Escherichia coli*. J. Bacteriol. 181:7005–7013.
132. Meisel, A., T. A. Bickle, D. H. Krüger, and C. Schroeder. 1992. Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. Nature 355:467–469.
133. Meisel, A., P. Mackeldanz, T. A. Bickle, D. H. Krüger, and C. Schroeder. 1995. Type III restriction endonucleases translocate DNA in a reaction driven by recognition site-specific ATP hydrolysis. EMBO J. 14:2958–2966.
134. Mise, K., and W. Arber. 1976. Plaque-forming transducing bacteriophage P1 derivatives and their behaviour in lysogenic conditions. Virology 69:191–205.
135. Moak, M., and I. J. Molineux. 2000. Role of the Gp16 lytic transglycosylase motif in bacteriophage T7 virions at the initiation of infection. Mol. Microbiol. 37:345–355.
136. Mukhopadhyay, G., K. M. Carr, J. M. Kaguni, and D. K. Chattoraj. 1993. Open-complex formation by the host initiator, DnaA, at the origin of P1 plasmid replication. EMBO J. 12:4547–4554.
137. Mukhopadhyay, G., S. Sozhamannan, and D. K. Chattoraj. 1994. Relaxation of replication control in chaperone-independent initiator mutants of plasmid P1. EMBO J. 13:2089–2096.
138. Mukhopadhyay, S., and D. K. Chattoraj. 2000. Replication-induced transcription of an autorepressed gene: the replication initiator gene of plasmid P1. Proc. Natl. Acad. Sci. USA 97:7142–7147.
139. Murray, N. E. 2000. Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). Microbiol. Mol. Biol. Rev. 64:412–434.
140. Nguyen, H. A., T. Tomita, M. Hirota, J. Kaneko, T. Hayashi, and Y. Kamio. 2001. DNA inversion in the tail fiber gene alters the host range specificity of carotovoricin Er, a phage-tail-like bacteriocin of phytopathogenic *Erwinia carotovora* subsp. *carotovora* Er. J. Bacteriol. 183:6274–6281.
141. Nordström, K., and S. J. Austin. 1989. Mechanisms that contribute to the stable segregation of plasmids. Annu. Rev. Genet. 23:37–69.
142. Park, K., and D. K. Chattoraj. 2001. DnaA boxes in the P1 plasmid origin: the effect of their position on the directionality of replication and plasmid copy number. J. Mol. Biol. 310:69–81.
143. Park, K., E. Han, J. Paulsson, and D. K. Chattoraj. 2001. Origin pairing ("handcuffing") as a mode of negative control of P1 plasmid copy number. EMBO J. 20:7323–7332.
144. Plasterk, R. H., and P. P. van de. 1985. The invertible P-DNA segment in the chromosome of *Escherichia coli*. EMBO J. 4:237–242.
145. Poteete, A. R., and L. W. Hardy. 1994. Genetic analysis of bacteriophage T4 lysozyme structure and function. J. Bacteriol. 176:6783–6788.
146. Redaschi, N., and T. A. Bickle . 1996. Posttranscriptional regulation of *Eco*P1I and *Eco*P15I restriction activity. J. Mol. Biol. 257:790–803.
147. Riedel, H. D., J. Heinrich, A. Heisig, T. Choli, and H. Schuster. 1993. The antirepressor of phage P1. Isolation and interaction with the C1 repressor of P1 and P7. FEBS Lett. 334:165–169.
148. Riedel, H. D., J. Heinrich, and H. Schuster. 1993. Cloning, expression, and characterization of the *icd* gene in the *imm*I operon of bacteriophage P1. J. Bacteriol. 175:2833–2838.
149. Rodionov, O., M. Lobocka, and M. B. Yarmolinsky. 1999. Silencing of genes flanking the P1 plasmid centromere. Science 283:546–549.
150. Rosner, J. L. 1972. Formation, induction, and curing of bacteriophage P1 lysogens. Virology 48:679–680.
151. Rydman, P. S., and D. H. Bamford. 2000. Bacteriophage PRD1 DNA entry uses a viral membrane-associated transglycosylase activity. Mol. Microbiol. 37:356–363.
152. Saha, S., and D. N. Rao. 1995. ATP hydrolysis is required for DNA cleavage by *Eco*PI restriction enzyme. J. Mol. Biol. 247:559–567.
153. Sandmeier, H., S. Iida, and W. Arber. 1992. DNA inversion regions Min of plasmid p15B and Cin of bacteriophage P1: evolution of bacteriophage tail fiber genes. J. Bacteriol. 174:3936–3944.
154. Sandmeier, H., S. Iida, P. Hübner, R. Hiestand-Nauer, and W. Arber. 1991. Gene organization in the multiple DNA inversion region min of plasmid p15B of *E. coli* 15T$^-$: assemblage of a variable gene. Nucleic Acids Res. 19:5831–5838.

155. Sandulache, R., P. Prehm, and D. Kamp. 1984. Cell wall receptor for bacteriophage Mu G(+). J. Bacteriol. 160:299–303.
156. Sauer, B. 1998. Inducible gene targeting in mice using the Cre/*lox* system. Methods in Enzymol. 14:381–392.
157. Schaefer, T. S., and J. B. Hays. 1990. The *bof* gene of bacteriophage P1: DNA sequence and evidence for roles in regulation of phage *c*1 and *ref* genes. J. Bacteriol. 172:3269–3277.
158. Schmidt, C., M. Velleman, and W. Arber. 1996. Three functions of bacteriophage P1 involved in cell lysis. J. Bacteriol. 178:1099–1104.
159. Schneider, T. D., and R. M. Stephens. 1990. Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. 18:6097–6100.
160. Sharrocks, A. D., and D. P. Hornby. 1991. Transcriptional analysis of the restriction and modification genes of bacteriophage P1. Mol. Microbiol. 5:685–694.
161. Shepherd, N. S., B. D. Pfrogner, J. N. Coulby, S. L. Ackerman, G. Vaidyanathan, R. H. Sauer, T. C. Balkenhol, and N. Sternberg. 1994. Preparation and screening of an arrayed human genomic library generated with the P1 cloning system. Proc. Natl. Acad. Sci. USA 91:2629–2633.
162. Shine, J., and L. Dalgarno. 1975. Terminal-sequence analysis of bacterial ribosomal RNA. Correlation between the 3′-terminal-polypyrimidine sequence of 16-S RNA and translational specificity of the ribosome. Eur. J. Biochem. 57:221–230.
163. Silverman, M., J. Zieg, M. Hilmen, and M. Simon. 1979. Phase variation in *Salmonella*: genetic analysis of a recombinational switch. Proc. Natl. Acad. Sci. USA 76:391–395.
164. Skorupski, K., J. C. Pierce, B. Sauer, and N. Sternberg. 1992. Bacteriophage P1 genes involved in the recognition and cleavage of the phage packaging site (*pac*). J. Mol. Biol. 223:977–989.
165. Skorupski, K., B. Sauer, and N. Sternberg. 1994. Faithful cleavage of the P1 packaging site (*pac*) requires two phage proteins, PacA and PacB, and two *Escherichia coli* proteins, IHF and HU. J. Mol. Biol. 243:268–282.
166. Skorupski, K., N. Sternberg, and B. Sauer. 1994. Purification and DNA-binding activity of the PacA subunit of the bacteriophage P1 pacase enzyme. J. Mol. Biol. 243:258–267.
167. Skowyra, D., and S. Wickner. 1993. The interplay of the GrpE heat shock protein and Mg^{2+} in RepA monomerization by DnaJ and DnaK. J. Biol. Chem. 268:25296–25301.
168. Sozhamannan, S., and D. K. Chattoraj. 1993. Heat shock proteins DnaJ, DnaK, and GrpE stimulate P1 plasmid replication by promoting initiator binding to the origin. J. Bacteriol. 175:3546–3555.
169. Sternberg, N. L. 1994. The P1 cloning system: past and future. Mamm. Genome 5:397–404.
170. Sternberg, N. L., and G. Cohen . 1989. Genetic analysis of the lytic replicon of bacteriophage P1. II. Organization of replicon elements. J. Mol. Biol. 207:111–133.
171. Sternberg, N. L., and J. Coulby. 1987. Recognition and cleavage of the bacteriophage P1 packaging site (*pac*). II. Functional limits of *pac* and location of *pac* cleavage termini. J. Mol. Biol. 194:469–479.
172. Sternberg, N. L., and J. Coulby. 1990. Cleavage of the bacteriophage P1 packaging site (*pac*) is regulated by adenine methylation. Proc. Natl. Acad. Sci. USA 87:8070–8074.
173. Sternberg, N. L., M. Powers, M. B. Yarmolinsky, and S. J. Austin. 1981. Group Y incompatibility and copy control of P1 prophage. Plasmid 5:138–149.
174. Sternberg, N. L., B. Sauer, R. Hoess, and K. Abremski. 1986. Bacteriophage P1 *cre* gene and its regulatory region. Evidence for multiple promoters and for regulation by DNA methylation. J. Mol. Biol. 187:197–212.
175. Streiff, M. B., S. Iida, and T. A. Bickle. 1987. Expression and proteolytic processing of the *darA* antirestriction gene product of bacteriophage P1. Virology 157:167–171.
176. Struthers-Schlinke, J. S., W. P. Robins, P. Kemp, and I. J. Molineux. 2000. The internal head protein Gp16 controls DNA ejection from the bacteriophage T7 virion. J. Mol. Biol. 301:35–45.
177. Summers, D., S. Yaish, J. Archer, and D. Sherratt. 1985. Multimer resolution systems of ColE1 and ColK: localisation of the crossover site. Mol. Gen. Genet. 201:334–338.
178. Surtees, J. A., and B. E. Funnell. 1999. P1 ParB domain structure includes two independent multimerization domains. J. Bacteriol. 181:5898–5908.
179. Surtees, J. A., and B. E. Funnell. 2001. The DNA binding domains of P1 ParB and the architecture of the P1 plasmid partition complex. J. Biol. Chem. 276:12385–12394.
180. Touati-Schwartz, D. 1979. A *dnaB* analog *ban*, specified by bacteriophage P1: genetic and physiological evidence for functional analogy and interactions between the two products. Mol. Gen. Genet. 174:173–188.
181. Touati-Schwartz, D. 1979. A new pleiotropic bacteriophage P1 mutation, *bof*, affecting *c*1 repression activity, the expression of plasmid incompatibility and the expression of certain constitutive prophage genes. Mol. Gen. Genet. 174:189–202.
182. Travers, A. 1997. DNA–protein interactions: IHF—the master bender. Curr. Biol. 7:R252–R254.
183. Treptow, N., R. Rosenfeld, and M. B. Yarmolinsky. 1994. Partition of nonreplicating DNA by the *par* system of bacteriophage P1. J. Bacteriol. 176:1782–1786.
184. Velleman, M., T. Heinzel, and H. Schuster. 1992. The Bof protein of bacteriophage P1 exerts its modulating function by formation of a ternary complex with operator DNA and C1 repressor. J. Biol. Chem. 267:12174–12181.
185. Velleman, M., M. Heirich, A. Günther, and H. Schuster. 1990. A bacteriophage P1-encoded modulator protein affects the P1 *c*1 repression system. J. Biol. Chem. 265:18511–18517.
186. Walker, D. H., and T. F., anderson. 1970. Morphological variants of coliphage P1. J. Virol. 5:765–782.
187. Walker, D. H., and J. T. Walker. 1983. Coliphage P1 morphogenesis: analysis of mutants by electron microscopy. J. Virol. 45:1118–1139.
188. Walker, D. H. Jr., and J. T. Walker. 1976. Genetic studies of coliphage P1. III. Extended genetic map. J. Virol. 20:177–187.
189. Wall, J. D., and P. D. Harriman. 1974. Phage P1 mutants with altered transducing abilities for *Escherichia coli*. Virology 59:532–544.

190. Wickner, S., J. Hoskins, D. Chattoraj, and K. McKenney. 1990. Deletion analysis of the mini-P1 plasmid origin of replication and the role of *Escherichia coli* DnaA protein. J. Biol. Chem. 265:11622–11627.
191. Wickner, S., D. Skowyra, J. Hoskins, and K. McKenney. 1992. DnaJ, DnaK, and GrpE heat shock proteins are required in *ori*P1 DNA replication solely at the RepA monomerization step. Proc. Natl. Acad. Sci. USA 89:10345–10349.
192. Williams, M. D., J. A. Fuchs, and M. C. Flickinger. 1991. Null mutation in the stringent starvation protein of *Escherichia coli* disrupts lytic development of bacteriophage P1. Gene 109:21–30.
193. Windle, B. E., and J. B. Hays. 1986. A phage P1 function that stimulates homologous recombination of the *Escherichia coli* chromosome. Proc. Natl. Acad. Sci. USA 83:3885–3889.
194. Windle, B. E., C. S. Laufer, and J. B. Hays. 1988. Sequence and deletion analysis of the recombination enhancement gene (*ref*) of bacteriophage P1: evidence for promoter–operator and attenuator–antiterminator control. J. Bacteriol. 170:4881–4889.
195. Yarmolinsky, M. 2000. Transcriptional silencing in bacteria. Curr. Opin. Microbiol. 3:138–143.
196. Yarmolinsky, M. B. 1995. Programmed cell death in bacterial populations. Science 267:836–837.
197. Yarmolinsky, M. B., and M. B. Lobocka. 1993. Bacteriophage P1, pp. 1.50-1.61. *In* S. J. O'Brian (ed.) Locus Maps of Complex Genomes. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
198. Yarmolinsky, M. B., and N. L. Sternberg. 1988. Bacteriophage P1, pp. 291–438. *In* R. Calendar (ed.) The Bacteriophages. Plenum Press, New York.
199. Young, M. C., S. B. Kuhl, and P. H. von Hippel. 1994. Kinetic theory of ATP-driven translocases on one-dimensional polymer lattices. J. Mol. Biol. 235:1436–1446.
200. Young, R. 1992. Bacteriophage lysis: mechanism and regulation. Microbiol. Rev. 56:430–481.

25

The P2-Like Bacteriophages

ANDERS S. NILSSON
ELISABETH HAGGÅRD LJUNGQUIST

Origin of Phage P2 and P2-Like phages

Bacteriophage P2 was isolated by G. Bertani, together with two other prophages, from the Lisbonne and Carrère strain of *Escherichia coli* (the oldest known lysogen) (8). The three phages were named P1, P2, and P3, based on different plaque morphology, and their partial serological crossreactivity. Phage P1 was later shown by E. Lennox to be a general transducing phage and to belong to a group of phages that differs from P2 in many respects (82) (see chapter 24). Phage P3 has not been studied as extensively as the other two. Since that time, a number of temperate phages that grow on *E. coli* have been isolated that share some, but not all, characteristics with phage P2. These P2-like phages are similar in traits such as morphology and host range, are serologically unrelated to phage λ and not inducible by UV light. Of these P2-like coliphages P2 and 186 are the best characterized, and they have in addition the capacity to function as helpers for phage/plasmid P4. For a list of other known P2-like *E. coli* phages, see reviews by Bertani and Bertani (11) and Bertani and Six (12), and for phage/plasmid P4 see chapter 26.

P2-Like Phages in E. coli and Other Bacteria

P2-like prophages seem to be quite common in *E. coli*. At least 26% of the strains in the *E. coli* reference collection (ECOR) contain a P2-like prophage: hybridization with a ^{32}P-labeled P2 DNA probe against chromosomal DNA resulted in a strong signal to 19 strains and a weak signal to two more strains of the collection (109). The 72 strains in the ECOR collection have been selected from a set of 2600 isolates from a variety of hosts and from different parts of the world (111). Thus, the ECOR collection is expected to contain a large part of the genetic variation in *E. coli*.

P2-like phages also seem to be distributed among other proteobacteria of the gamma subgroup. The genomes of phages HP1 and HP2 in *Haemophilus influenzae* (35, 151), phage ΦCTX in *Pseudomonas aeruginosa* (103), and phage K139 in *Vibrio cholerae* (105) have all been sequenced and shown to be P2-like with respect to genome organization as well as nucleotide sequence. Phages PSP3 in *Salmonella potsdam* (20) and SopEΦ in *Salmonella typhimurium* (100) are also P2-like, but they are not yet fully sequenced. Bacterial genome sequencing projects have revealed additional P2-like prophages, for example Sp13 in the enterohemorrhagic *E. coli* O157:H7 isolated during the Sakai outbreak in Japan (55), and Fels-2 from *Salmonella typhimurium* LT2 (99).

Scope of this Chapter

Since the last comprehensive review on phage P2 and P2-like phages, written in 1988 (12), new information concerning many aspects of P2's gene regulation, DNA replication, head morphogenesis, and site-specific recombination has accumulated. The current understanding of these topics will be summarized here. Since several P2-like phages have been completely sequenced recently, it has become possible to compare whole genomes and shed some light on the evolution of P2-like phages. But this has also revealed large differences, which makes classification of phages as belonging to a specific "family" or "group" problematic. In some cases, phages have similar genes encoding capsid proteins, while their DNA replication machinery seems to differ, for example phage P2 compared with phage ΦCTX. In other cases, phages share replication mechanism with phage P2, like the virulent phage ΦX174 and phage PM2, but nothing more. It is too early to decide which phage characters should be used to define a specific family since too few P2-like phages have been fully sequenced. There functional gene groups are compared rather than whole phages, and we use the term "P2-like" for phages that share some, but not all, properties with P2.

The genetic interactions between phage P2 and the defective satellite phage, or plasmid, P4 is discussed in detail in chapter 26, and only certain key features relevant for the P2 helper are discussed in this chapter.

Phage Structure and Assembly

A P2-like phage has an icosahedral head with a diameter of about 60 nm, containing a linear double-stranded DNA molecule of about 30–35 kb with cohesive ends 19 nucleotides long, and a 135 nm long straight tail with a contractile sheath. The tail ends with a baseplate carrying six tail fibers and a spike. Based on their morphology, P2-like phages are taxonomically classified, together with phage P1 and the T-even phages, as members of the *Myoviridae* family in the order *Caudovirales* (1).

The Capsid

The major head protein of phage P2 is encoded by gene *N*, and the mature capsid contains mainly the processed form of the N protein, N*, which lacks 31 amino acids at the N-terminus (126). The structure of the P2 head has been analyzed by image reconstructions from cryo-electron micrographs with a resolution of 4.5 nm (32). The N protein seems to have two domains, one domain comprising the capsomer and the other forming trimeric connections between the capsomers. The capsomers are assembled in a T = 7 icosahedral symmetry with 12 pentamers and 60 hexamers. The pentamers protrude more from the capsid surface than the hexamers.

P2 head assembly has been extensively studied, since the capsid protein N can assemble into two different sizes, 60 or 45 nm, depending on the scaffolding protein used (2, 97, 98). In a normal P2 infection, only large capsids are formed, but during a mixed infection with satellite phage P4, small capsids with a T = 4 symmetry are also formed. The small capsids can only package the small 11.52 kb P4 genome, thereby excluding the large 33.6 kb P2 genome. Cryo-electron microscopy studies of procapsid-like structures isolated from cells expressing the N protein have shown that closed shells of both sizes are formed, although with a low efficiency. Coproduction of P2 N protein and the P2 scaffolding protein, O, leads to a more efficient assembly of large capsid structures, but when the phage P4 scaffolding protein, Sid, is also expressed, the assembly of small capsid structures prevails (98). The phage P2 O protein is presumed to form an internal scaffold, while the P4 Sid protein forms an external scaffold (31, 96). The P2 O and the P4 Sid proteins are both required for formation of infectious P4 particles. The target for the phage P4 scaffolding protein, Sid, has been identified by the isolation of phage P2 mutants (P2 *sir* mutants for sid responsiveness) that do not respond to the action of the P4 Sid protein. The *sir* mutations have been mapped to a 38 codon long segment in the middle of gene *N* (137).

The procapsid contains the full-length N protein and the scaffold protein, O, which implies that processing of the N and O proteins occurs after assembly. Thus, the mature capsid contains the processed form of N protein, called N*, as well as the 17 kDa processed form of the O protein (97, 126).

The products of genes *Q*, *P*, and *M* are required for packaging the DNA into the head and for conversion of the prohead to a capsid. The Q protein constitutes the portal protein (see below), while the M and P proteins constitute the terminase activity. The packaging substrate consists of closed monomeric DNA circles that are cleaved at the *cos* site generating the 19-base long single-stranded cohesive ends. The M protein, which has a DNA-binding activity, has been suggested to contain the endonuclease activity, while the P protein contains a DNA-dependent ATPase activity that can account for the ATP requirement of the terminase activity (16, 17).

An alignment of the *cos* regions of phages P2 and P4 identified a region of 55 bp, including the common 19 bp long *cos* sequence, with only three mismatches. A conserved inverted repeat, located at one side of the *cos* sequences, might be the recognition sequence for the packaging protein (161).

The Connector

The connector or portal that joins the phage head and tail has a double-disk structure. One disk is hidden by the head and is detectable only after disruption of the virions (127). The phage P2 gene *Q* encodes the connector protein, which is present in the virion in a processed form that lacks the 24 N-terminal amino acids (86, 127). Purified full-length Q protein assembles into connector-like structures. Image reconstructions, based on two-dimensional crystalline layers of purified full-length Q proteins, have shown a toroid structure with a central mass surrounding a channel with a diameter of about 2 nm. The central mass has 12 protrusions that suggest a 12-fold symmetry (125). The P2 connector thus seems to have a design similar to other phage connectors even though no similarities at the DNA or protein levels have been detected (86, 95, 145, 146).

The Tail

There are 12 genes known to be involved in tail production in the P2 genome, of which seven have been identified as part of the tail structure (81, 84).

Proteins F_I and F_{II} make up 60% and 30% of the tail respectively, and have been proposed to constitute the sheath and the tube respectively (81, 142).

Mutations in tail genes *R* and *S* will still result in tail structures. Infections with phage P2 *R* mutants result in giant naked tail tubes and extended tails, while the tails of the *S* mutants look normal but are functionally inactive (81). Due to these phenotypes, the S and R proteins have been suggested to be involved in tail completion (85). The T protein is present in the tail structure with 5 ± 2 copies per tail. Its function is not known, but due to its large size and predicted α-helical structure, it has been suggested to be the ruler protein (85).

The tail is terminated by a baseplate, which carries six tail fibers and a small protruding spike. Electron microscopy of phage particles, using antibodies raised against proteins V or J, has shown that the V protein forms the spike and that the J protein is part of the baseplate or the proximal part of the tail fiber (48). The W protein has been inferred to be part of the base plate since it is homologous to the gene 25 protein of phage T4. The H protein probably constitutes the distal part of the tail fiber since it contains regions similar to tail fiber proteins of other unrelated phages with the same host range (47). The assembly of the tail fiber is probably mediated by the G protein.

Comparison with Other P2-Like Phages

The structural genes of P2-like phages are arranged similarly on the genetic map and many of the encoded proteins are very similar in composition and size (116, 153) (table 25-1). Clearly the genes required for capsid formation are very well conserved in all six fully sequenced P2-like phages, while the tail genes fall into two groups based on sequence similarities and gene organization. One group contains phages P2, 186, and ΦCTX, and the other contains phages HP1, HP2, and K139. A possible explanation for the difference in tail gene organization is that all HP1 tail genes are transcribed from a single promoter, and genes encoding proteins required in large quantities must be located close to the promoter (35). In general, P2-like tail genes are found in both related and unrelated phages, as well as in bacteria where they can function as bacteriocins (35, 92, 104, 115).

The Lytic Cycle

P2-like phages are temperate, that is they can grow lytically as well as forming lysogens. During lytic growth, gene expression is regulated over time. Early transcription is initiated immediately after infection, and requires only the host σ^{70} RNA polymerase, leading to expression of the genes required for DNA replication. Once DNA replication has been initiated and the transcriptional activator needed for activation of the late promoter has accumulated, late gene transcription is initiated. The capsid and tail are assembled by two independent pathways. After DNA packaging the tails are added to the filled capsids. As is true for all tailed phages, the phage P2 lytic cycle ends with phage-induced cell lysis (chapter 10).

Early Transcription

The early operons of phages P2, 186, HP1, and K139 contain 9, 12, 11, and 11 genes, respectively (figure 25-1). The first gene of each operon, designated *cox* in P2, HP1, and K139, but *apl* in phage 186, encodes the repressor of the lysogenic promoter. These genes are thus functionally equivalent to λ *cro*, but unlike the λ Cro potein, P2 Cox and 186 Apl have been shown not to be essential for lytic growth even though at high concentrations they negatively reduce early lytic transcription (119, 129). Another common gene in the early transcript is the *A* gene, designated *rep* in HP1 and K139, that is required for initiation of rolling circle replication (see DNA Replication below). It should be noted that phage ΦCTX does not contain a gene homologous to the P2 *A* gene and may thus initiate DNA replication by some other mechanism. For phage 186, and possibly phage HP1, the initiator is the only phage protein required for phage DNA replication (136). This is in contrast to phage P2 where gene *B* encodes an additional protein, a helicase loader (112). Phage K139 may also encode a P2 B-like protein, since the product of ORF3 has 28% identity to P2 B (72). The early transcript of phage 186 contains a *cII* gene, which encodes a transcriptional activator of the phage 186 P_E promoter required for establishment of lysogeny (77, 106). The phage 186 P_E promoter, located between genes *apl* and *cII*, controls expression of gene *cI*. Phage K139 has a gene encoding a protein showing 27% identity to phage 186 cII protein and it is therefore believed to have a similar function (105). Most of the remaining genes have unknown functions, and some—phage P2 ORF78, ORF80, and ORF82, and phage 186 *dhr* and *fil* genes (92, 123, 129)—are lethal to the host. The *dhr* and *fil* gene products have been shown to inhibit host DNA replication and cell division, respectively. Phage P2 ORF82 and ORF83 are similar to phage 186 ORF80 and ORF83, but in phage 186 they are separated by ORF81. Proteins with 30–40% identity to that produced by the P2 ORF82 can be found in plasmids (*E. coli* F plasmids), and in phages (ΦCTX, 933W), and in bacteria (*E. coli*, *H. influenzae*, *Yersinia pestis*, and *Treponema pallidum*), and it has been suggested to be a *dnaK* suppressor (33).

Early lytic transcription has been analyzed only in phage 186. It is terminated at *t*R1, a Rho-independent terminator that shows 70% efficiency in vitro (124). The transcript is extended by some unknown antitermination mechanism that does not require any known phage 186 protein. The extended transcript is processed by RNaseIII which cleaves at a site within the *fil-dhr* region (28).

Many of the genes in the early operons of phage P2 and phage 186 have start and stop codons that overlap, indicating translational coupling. This has also been shown in phage 186, where an amber mutation in ORF84, located proximal to the *A* gene, prevents expression of the A initiator protein, which lacks a ribosomal binding site (136). In the case of phage P2, ORF83 is located proximal to the *A* gene, and the two genes overlap by three amino acids. Since phage P2 gene *A* also seems to lack a properly spaced ribosome-binding site, the expression of A is very low (80; A. Ahlgren Berg, personal communication).

Table 25-1 Late Genes of P2-Like Phages

P2		186		ΦCTX		HP1/HP2		K139			
Gene	Codons	Gene	Codons	Gene	Codons	Gene	Codons	Gene	Codons	Similar genes	Function
				ctx	286						Toxin
Q	344	*Orf2*	340	*Q*	350	*Orf15*	345/345	*Orf15*	348	Ec67 *Orf5*	Capsid portal protein
–		*W*	248								
P	590	*Orf12*	589	*P*	594	*Orf16*	607/607	*Orf16*	605	Ec67 *Orf6*	Large terminase subunit
O	284	*V*	284	*O*	273	*Orf17*	298/297	*Orf17*	299	ΦHs *Orf1*	Capsid scaffold
N	357	*T*	355	*N*	338	*Orf18*	336/336	*Orf18*	341	ΦHs *Orf2*	Major capsid precursor
M	247	*R*	249	*M*	235	*Orf19*	281/279	*Orf19*	238	ΦHs *Orf3*	Small terminase subunit
L	169	*Q*	168	*L*	153	*Orf20*	150/150	*Orf20*	153	ΦHs *Orf4*	Capsid completion
				Orf7	65						
X	67	*Orf23*	67	*X*	69					Pyocin R *Orf22* ΦHs *Orf5*	Tail
Y	93	*Orf24*	98								Lysis — holin
						Orf21	166/161	*Orf21*	162		
								Orf23	53		
						Orf23	376/376	*Orf24*	369		Tail sheath
						Orf24	150/144	*Orf25*	152		Tail tube
								Orf26	69		P2 Orf82
								Orf27	75		
						hol	78/78			ΦHs *Orf6*	Lysis — holin
K	165	P	165			*lys*	186/179	*Orf28*	195	ΦHs *Orf7*, λ *R*	Lysis — endolysin
lysA	141										Lysis — timing
				Orf9	117						
				Orf10	90						
				Orf11	268						
lysB	141	*Orf27*	137	*lysB*	153						Lysis — timing
Orf		*Orf28*	96	*Orf12.5*	89						
R	155	*N*	155	*R*	178					ΦHs *Orf10*	Tail completion
S	150	*O*	149	*S*	156	*Orf22*	227/231	*Orf22*	166	ΦHs *Orf11*	Tail completion
				Orf15	242						
Orf30	261										
V	211	*Orf32*	213	*V*	190					Pyocin R *Orf11*	Tail spike
W	115	*M*	115	*W*	114					Pyocin R *Orf12*	Baseplate
J	302	*L*	302	*J*	304					Pyocin R *Orf13*	Baseplate/tail fiber
I	176	*Orf38*	176	*I*	178					Pyocin R *Orf14*	Tail
						Orf25	115/115				
								Orf29	116		Signal peptide
						Orf26	102/102	*Orf30*	192		Signal peptide
						Orf28	111/111	*Orf32*	110		
						Orf29	393/382	*Orf33*	399		
						Orf30	174/175	*Orf34*	219		
H	669	*K*	462	*H*	762	*Orf31*	925/910	*Orf35*	620	Pyocin R *Orf15*	Tail fiber
G	175	*Orf45*	166			*Orf32*	200/210			Mu gene *U*	Tail fiber assembly
Z/fun	528										Lysogenic conversion
								Orf36	174		Tail fiber assembly
				Orf21	148						
FI	396	*J*	392	*FI*	391					Pyocin R *Orf17*	Tail sheath
FII	172	*I*	173	*FII*	171					Pyocin R *Orf18*	Tail tube
E+E	142	*H*	162	*E*	108						
E	91	*Orf52*	58	*E′*	39						
T	815	*G*	812	*T*	904	*Orf27*	689/709	*Orf31*	605		Tail
U	159	*F*	161	*U*	146					Pyocin R *Orf21*	Tail
D	387	*D*	389	*D*	424					Pyocin R *Orf23*	Tail
				Orf28	295						
				Orf29	170						
				Orf30	116						
				Orf31	38						
				Orf32	60						
				Orf33	156						
						Orf33	258/255	*Orf37*	297		
						Orf34	187/186	*Orf38*	157		
						Orf35	533/533	*Orf39*	543		
ogr	72	*B*	72	*ogr*	97						Late promoter activator

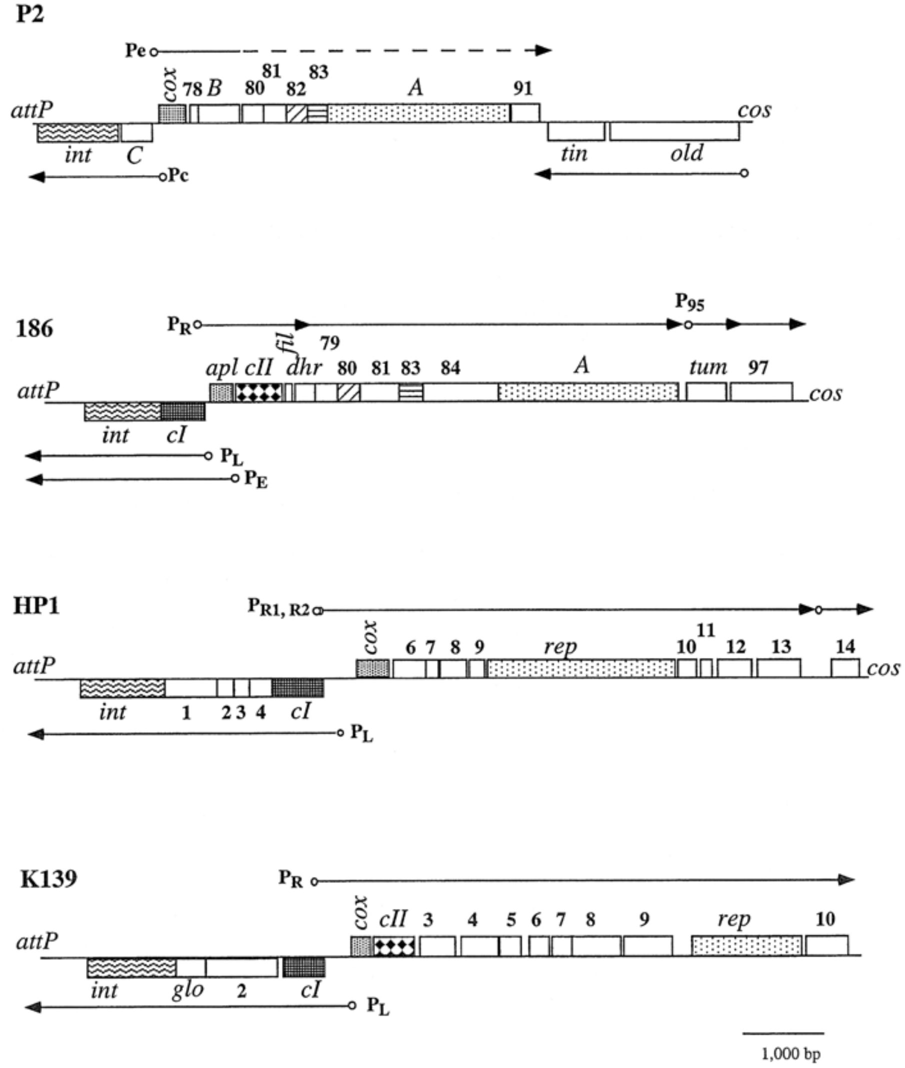
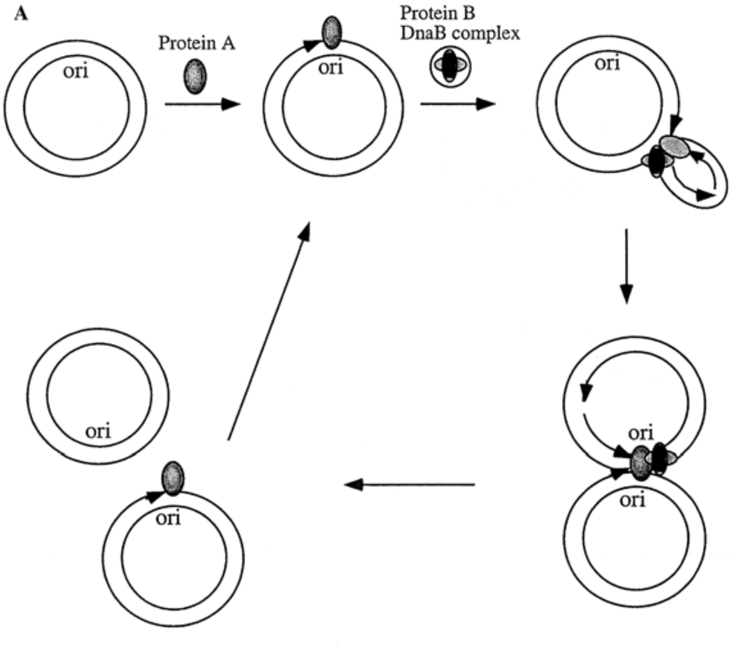


Figure 25-1 Schematic drawing of the regions between *attP* and *cos*, or the equivalent region, of phage P2 and some P2-like phages. The arrows indicate direction of transcription, and the small circle the respective promoter. The genes, or open reading frames (ORFs), are indicated above or below the genes and coding parts are indicated by boxes. Homologous genes, or ORFs, are indicated by identical fillings of the respective box.

DNA Replication

DNA replication has been studied only in phages P2 and 186, which both replicate via a modified rolling-circle mechanism that generates double-stranded monomeric circles. DNA replication is initiated by the A protein, which catalyzes a single-stranded cut at the origin. Upon cleavage, the A protein gets covalently linked to the 5′ end of the cleaved strand, while the 3′-OH end will function as a primer for DNA polymerization (112). After one round of replication, when the origin sequence is regenerated, the covalently linked A protein initiates a series of transesterification reactions which lead to cleavage of the newly synthesized origin sequence and, at the same time, joining of the displaced old strand (figure 25-2). In this way, the A protein becomes covalently linked to the 5′ end of the new strand and another cycle is directly initiated. In phage P2, the cleavage and joining reactions are mediated by two tyrosine residues located at the catalytic site of the A protein, Tyr-450 and Tyr-454 (90, 112). This DNA replication mechanism is thus very similar to the well-studied ΦX174 system, but in contrast to the ΦX174 system, the two tyrosine residues do not have equivalent roles during initiation of replication. The Tyr-454 residue of the phage P2 A protein promotes the initial cleavage reaction, after which the two tyrosine residues act in a flip-flop mechanism as has been suggested for phage ΦX174 (149).

The origin has been shown to be located within the coding part of the phage P2 *A* gene (91), and similar sequences can be found within the coding part of initiator proteins of phages 186, HP1, HP2, and K139 (figure 25-2). The importance of the bases at the cleavage site in the phage P2 origin



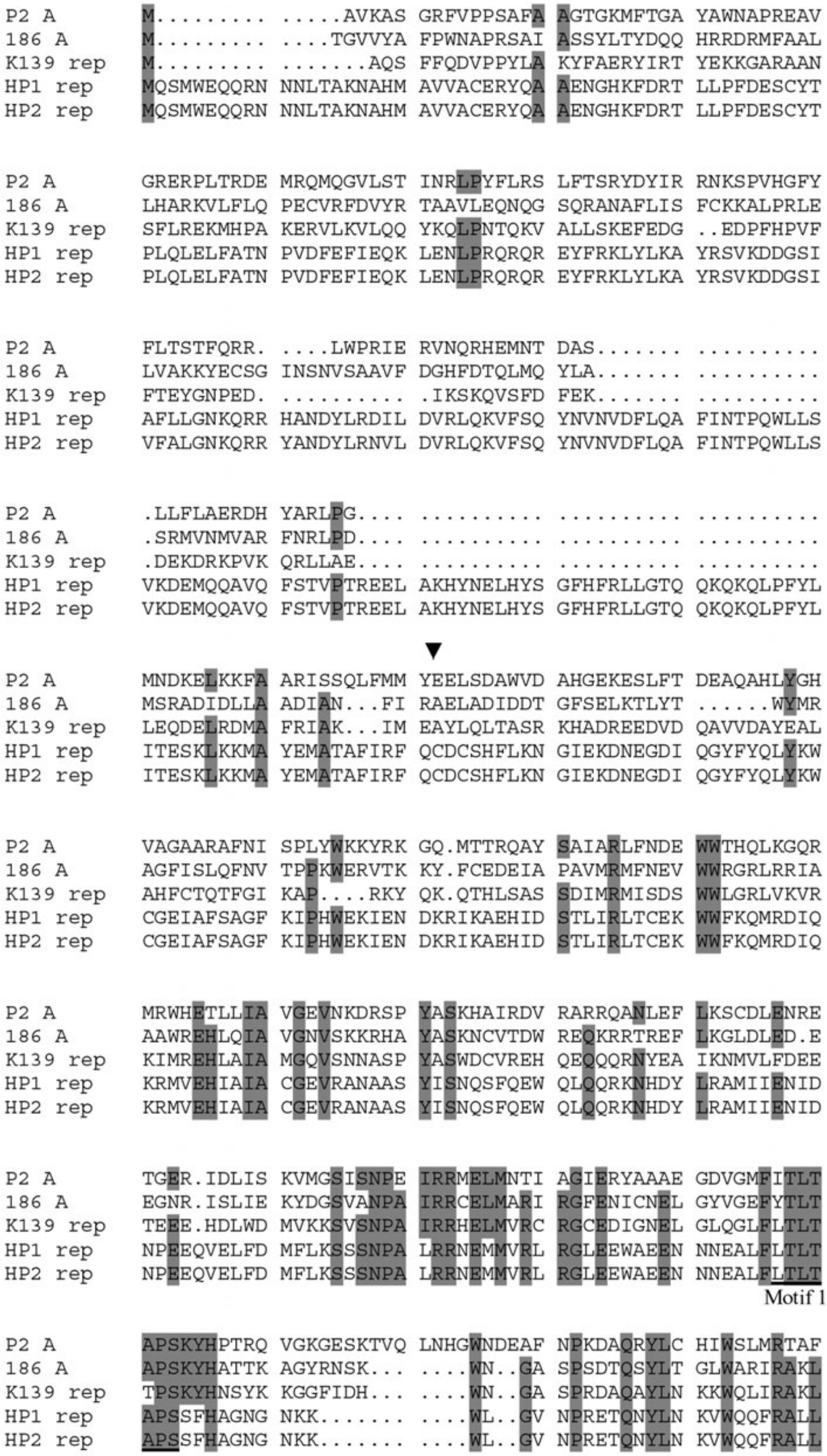
B

```
P2          CGCCGCGCCTCG/GAGTCCTGTCAATAACTGTGGAA
186         TGCGCCCTCTCG/GAGTTCTGTCAATAACTGTACGG
K139        TGCCGCCTCTCG/GAGTTCTGTCAATAACTGTACGG
HP1/HP2     CAGTGCGCCTTG/GACTTGTGTCAGTAACTGTAACC
Consensus    gC gC  CTcG/GAgTtcTGTCAaTAACTGTa
```

Figure 25-2 Phage P2 DNA replication. A: Hypothetical model of the steps during P2 DNA replication. Initiation: The A protein is synthesized, cleaves the *ori*-region in *cis*, and gets covalently linked to the 5′ end of the cleaved strand. The A protein recruits the P2 B/DnaB helicase complex, REP, and all other host factors required for DNA replication to the origin. Elongation: The free 3′-OH end is used as primer for the DNA polymerase III and the displacement of the old strand, which contains the A protein at the 5′ end, is mediated by the REP helicase. The DnaB protein is required for lagging strand synthesis, possibly for recruitment of the DnaG primase. Termination: After one round of replication, when *ori* has been regenerated, the A protein cleaves the new *ori* and joins the old strand, which is released as a double stranded covalent-closed circle after lagging strand is finished. Since the A protein remains covalently linked to the 5′ end of the new strand, a new round of replication can be initiated. B: A comparison of the P2 origin of replication with the hypothesized origins of some P2-like phages. The cleavage site in P2 origin is indicated by a slash, and it has been determined experimentally, while the others are hypothesized. The nucleotides required for cleavage in P2 are in bold typeface. In the consensus sequence, nucleotides common to all five phages are capitalized.

has been analyzed, and two nucleotides are critical for in vitro cleavage: G at position −1 and T at position −3 (90). These two nucleotides are also conserved in the presumed origins of phages 186, K139, and HP1. Since the origin is located within the coding part of the *A* gene, this conservation might simply reflect essential amino acids in the A protein. However, the conserved T residue is located in the third position of the codon, and codes for a Pro in the initiation proteins of phages P2, HP1, and HP2 but a Ser in phages 186 and K139. Furthermore, the G residue is located in the second position and codes for an Arg in phages P2, 186, and K139 but Try in phages HP1 and HP2 (figure 25-3). The sequence on the 5′ side of the cleavage site is better conserved than that on the 3′ side, and might constitute the recognition sequence for the initiator proteins. The recognition sequence of the ΦX174 A protein is located at a similar position, but there is no significant sequence identity of this region to equivalent regions in the P2-like phages (4).

The A/Rep proteins of phages P2, 186, HP1, HP2, and K139 belong to a large family of proteins that initiate rolling-circle replication. This mode of replication is found among small phages, such as ΦX174 and M13, in a number of small plasmids in bacteria as well as in archaebacteria, and in plant and animal viruses. Conjugative plasmids have also been found to use a similar initiation mechanism for mobilization of their DNA. A comparative analysis of the initiation proteins has classified them into two major groups—the replication (Rep) and the mobilization (Mob)



```
P2 A       M......... .....AVKAS GRFVPPSAFA AGTGKMFTGA YAWNAPREAV
186 A      M......... ....TGVVYA FPWNAPRSAI ASSYLTYDQQ HRRDRMFAAL
K139 rep   M......... .......AQS FFQDVPPYLA KYFAERYIRT YEKKGARAAN
HP1 rep    MQSMWEQQRN NNLTAKNAHM AVVACERYQA AENGHKFDRT LLPFDESCYT
HP2 rep    MQSMWEQQRN NNLTAKNAHM AVVACERYQA AENGHKFDRT LLPFDESCYT

P2 A       GRERPLTRDE MRQMQGVLST INRLPYFLRS LFTSRYDYIR RNKSPVHGFY
186 A      LHARKVLFLQ PECVRFDVYR TAAVLEQNQG SQRANAFLIS FCKKALPRLE
K139 rep   SFLREKMHPA KERVLKVLQQ YKQLPNTQKV ALLSKEFEDG ..EDPFHPVF
HP1 rep    PLQLELFATN PVDFEFIEQK LENLPRQRQR EYFRKLYLKA YRSVKDDGSI
HP2 rep    PLQLELFATN PVDFEFIEQK LENLPRQRQR EYFRKLYLKA YRSVKDDGSI

P2 A       FLTSTFQRR. ....LWPRIE RVNQRHEMNT DAS....... ..........
186 A      LVAKKYECSG INSNVSAAVF DGHFDTQLMQ YLA....... ..........
K139 rep   FTEYGNPED. .......... .IKSKQVSFD FEK....... ..........
HP1 rep    AFLLGNKQRR HANDYLRDIL DVRLQKVFSQ YNVNVDFLQA FINTPQWLLS
HP2 rep    VFALGNKQRR YANDYLRNVL DVRLQKVFSQ YNVNVDFLQA FINTPQWLLS

P2 A       .LLFLAERDH YARLPG.... .......... .......... ..........
186 A      .SRMVNMVAR FNRLPD.... .......... .......... ..........
K139 rep   .DEKDRKPVK QRLLAE.... .......... .......... ..........
HP1 rep    VKDEMQQAVQ FSTVPTREEL AKHYNELHYS GFHFRLLGTQ QKQKQLPFYL
HP2 rep    VKDEMQQAVQ FSTVPTREEL AKHYNELHYS GFHFRLLGTQ QKQKQLPFYL

                                  ▼
P2 A       MNDKELKKFA ARISSQLFMM YEELSDAWVD AHGEKESLFT DEAQAHLYGH
186 A      MSRADIDLLA ADIAN...FI RAELADIDDT GFSELKTLYT ......WYMR
K139 rep   LEQDELRDMA FRIAK...IM EAYLQLTASR KHADREEDVD QAVVDAYEAL
HP1 rep    ITESKLKKMA YEMATAFIRF QCDCSHFLKN GIEKDNEGDI QGYFYQLYKW
HP2 rep    ITESKLKKMA YEMATAFIRF QCDCSHFLKN GIEKDNEGDI QGYFYQLYKW

P2 A       VAGAARAFNI SPLYWKKYRK GQ.MTTRQAY SAIARLFNDE WWTHQLKGQR
186 A      AGFISLQFNV TPPKWERVTK KY.FCEDEIA PAVMRMFNEV WWRGRLRRIA
K139 rep   AHFCTQTFGI KAP....RKY QK.QTHLSAS SDIMRMISDS WWLGRLVKVR
HP1 rep    CGEIAFSAGF KIPHWEKIEN DKRIKAEHID STLIRLTCEK WWFKQMRDIQ
HP2 rep    CGEIAFSAGF KIPHWEKIEN DKRIKAEHID STLIRLTCEK WWFKQMRDIQ

P2 A       MRWHETLLIA VGEVNKDRSP YASKHAIRDV RARRQANLEF LKSCDLENRE
186 A      AAWREHLQIA VGNVSKKRHA YASKNCVTDW REQKRRTREF LKGLDLED.E
K139 rep   KIMREHLAIA MGQVSNNASP YASWDCVREH QEQQQRNYEA IKNMVLFDEE
HP1 rep    KRMVEHIAIA CGEVRANAAS YISNQSFQEW QLQQRKNHDY LRAMIIENID
HP2 rep    KRMVEHIAIA CGEVRANAAS YISNQSFQEW QLQQRKNHDY LRAMIIENID

P2 A       TGER.IDLIS KVMGSISNPE IRRMELMNTI AGIERYAAAE GDVGMFITLT
186 A      EGNR.ISLIE KYDGSVANPA IRRCELMARI RGFENICNEL GYVGEFYTLT
K139 rep   TEEE.HDLWD MVKKSVSNPA IRRHELMVRC RGCEDIGNEL GLQGLFLTLT
HP1 rep    NPEEQVELFD MFLKSSSNPA LRRNEMMVRL RGLEEWAEEN NNEALFLTLT
HP2 rep    NPEEQVELFD MFLKSSSNPA LRRNEMMVRL RGLEEWAEEN NNEALFLTLT
                                                          Motif 1

P2 A       APSKYHPTRQ VGKGESKTVQ LNHGWNDEAF NPKDAQRYLC HIWSLMRTAF
186 A      APSKYHATTK AGYRNSK... ....WN..GA SPSDTQSYLT GLWARIRAKL
K139 rep   TPSKYHNSYK KGGFIDH... ....WN..GA SPRDAQAYLN KKWQLIRAKL
HP1 rep    APSSFHAGNG NKK....... ....WL..GV NPRETQNYLN KVWQQFRALL
HP2 rep    APSSFHAGNG NKK....... ....WL..GV NPRETQNYLN KVWQQFRALL
```

Figure 25-3 Alignment of the initiation proteins of P2 and some P2-like phages. The amino acids were aligned using the CLUSTAL X program (59). Amino acids identical in at least four proteins are indicated by dark gray shading. The common motifs for proteins initiating rolling-circle replication are underlined (62, 75). The location of P2 *ori* (in the corresponding DNA sequence), and the amino acid in phage 186 believed to interact with the REP protein, are indicated by arrowheads.

```
P2 A       KDNDLQVYGL RVVEPHHDGT PHWHMMLFCN PRQRNQIIEI MRRYALKEDG
186 A      HREEIRIFGI RVAEPHHDGT PHWHMLMFML PEDVERVRLI IRDYAWEEDH
K139 rep   NRDEIRWFGV RVAEPHHDGT PHWHLLIWVR KEDISAVRDT FITYATEEDR
HP1 rep    SKRNIKFYGM RVAEPHKDGT PHWHALAYVP AEHKEEVIRL FKQKALELDG
HP2 rep    SKRNIKFYGM RVAEPHKDGT PHWHALAYVP AEHKEEVIRL FKQKALELDG
                              Motif 2

P2 A       DERG...... .......... ....AARNRF QAKHLN..QG GAAGYIAKYI
186 A      YELRSDK... .......... ....AKKARF HAEAIDPEKG SATGYVAKYI
K139 rep   GELHPEFEKE KQKPFRKGVY VGPLDYRPRC DFGYIDPTKG TATGYIAKYI
HP1 rep    NEKG...... .......... ....AAEHRC KVEKCDKTKG SATAYIAKYI
HP2 rep    NEKG...... .......... ....AAEHRC KVEKCDKTKG SATAYIAKYI
                                                         Motif 3

P2 A       SKNIDGYALD GQLDND.TGR PLKDTAAAVT AWASTWRIPQ FKTVGLPTMG
186 A      SKNIDGYALD GETDDE.SGE LLKETAPAVS AWAARWHIRQ FQFIGGAPVT
K139 rep   SKNIDGYAMD GDISDE.TGK PVKDMARNVS AWKSRWSIRQ FQFFGGAPVT
HP1 rep    AKNIDGFALA GEVSDEDPTL SLHDNALRVR AWASRWGIRQ FQFYGGASIC
HP2 rep    AKNIDGFALA GEVSDEDPTL SLHDNALRVR AWASRWGIRQ FQFYGGASIC


P2 A       AYRELRKLPR G......... .......... .......... ..........
186 A      VYRELRRMAD .......... .......... .......... ..........
K139 rep   TYRELRRLAN QNKKAFMEYI FMQERADLIS MYELLHYQLI GAFKPARVMT
HP1 rep    VWRELRRLIS G......... .......... .......... ..........
HP2 rep    VWRELRRLIS G......... .......... .......... ..........

P2 A       .......... ..VSIADEFD ERVEAARAAA DSGDFALYIS AQGGANVPRD
186 A      .......... ..PETARALS VEFAAVHDAA HYGRWADYVN AQGGPFVRRD
K139 rep   NQELVEVIAQ SYEARAKTEI PHVAAVLRSA DEGRWHGYIM NQGGPFVKRK
HP1 rep    .......... ...QADDEII NKAQAAAGIA N..DYAAYME IQGGALAKRA
HP2 rep    .......... ...QADDEII NKAQAAAGIA N..DYAAYME IQGGALAKRA

P2 A       CQTVRVARSP SDEVNEYEEE VERVVGIYAP HLGARHIHIT RTTDWR....
186 A      DLQVRTLYEP RTEFNQYGEE TVCIKGVYDA SIGAGSPILT RLTQWK....
K139 rep   ELLVTNVYQE LPFASPYAEA IRKLEGIATP EQVIKTREKV WTIKRKGKET
HP1 rep    DQPIKLDYET KP.ANKYGEQ RKAIIGLANR FSLKQVISRT KKWQIKKRPQ
HP2 rep    DQPIKLDYET KP.ANKYGEQ RKAIIGLANR FSLKQVISRT KKWQIKKRPQ

                                     ori
                                     ▼
P2 A       .IVPKVPVVE PLTLKS...G IAAPRSPVNN CGKLTGGDTS LPAPTPSEHA
186 A      .IVPKRAVDL AVDVKG...A SAPSRSSVNN C......... ....TGSESD
K139 rep   TEGEAVGFGS EATAFG...G SAASRSSVNN CTDP...... ...FTGQVST
HP1 rep    DFAQRTESMV ERSSTANNSA RSAPWTCVSN CN........ .....RSILE
HP2 rep    DFAQRTESMV ERSSTANNSA RSAPWTCVSN CN........ .....RSNLE

P2 A       AAVLNLVDDG VIEWNEPEVV RALRGALKYD MRTPNRQQRN GSPLKPHEIA
186 A      PPILDLTKP. ....LSRRER RELTNRLRKK KPTTRRKFIH GTDKQNVAIT
K139 rep   QLTRLLQPDR LKGSQNDQIV DEVAISALFK GSTLRLDDET ELKIRPAEVD
HP1 rep    QKIKLLTQP. ...ICAPLSA QKLDYLFKYK RLTIDKYTAL ELTENDVQLV
HP2 rep    QKIKLLTQP. ...ICTPLSA QKLDYLFKYK RLTIDKYTAL ELTENDVQLV

P2 A       PSARLTRSER LQITRIRVDL AQNGIRPQRW ELEALARGAT VNYDGKKFTY
186 A      KTIDEIHSDN RHHNQPGRSP APDGRW
K139 rep   EHGKIRPARL VEVKREVDSD IWCRFEGWEK FEQQMEKLNQ KPQEHSQPDL
HP1 rep    KRNQNMMTSL SPVSRNFQKL KDFHKNQRIQ
HP2 rep    KRNQNMMTSL SPVPRNLQKL KNFHKKQRIQ

P2 A       PVADEWPGFS TVMEWT
K139 rep   SFFQELEGDW PLA...
```

Figure 25-3 Continued.

group—based on the arrangement of three conserved motifs (62, 75). The A/Rep proteins of these P2-like phages belong to the Rep class of initiation proteins. This class has the conserved motif 1, of unknown function, closest to the N-terminus. Motif 2 is located in the center and it contains two invariant His residues that are believed be involved in coordination of Mg^{2+} or Mn^{2+}. Motif 3, which contains the catalytic site with the two conserved tyrosine residues, is located closest to the C-terminus. A comparison of the initiation proteins of the P2-like phages is shown in figure 25-3 (72, 92, 136). Apart from the regions around the three conserved motifs, the initiation proteins share few sequence identities, but the Rep proteins of phages HP1 and HP2 are almost identical. It has been shown that up to 101 bp can be removed from the C-terminus of the phage P2 A protein without any significant effects on its biological activity (112). In fact as much as 150 bp can be removed, and the protein will still be able to cleave and join single-stranded substrates, although with a reduced efficiency (T. Krokeide, B. H. Lindqvist and E. Haggård-Ljungquist, unpublished data).

The N-terminal parts of the initiation proteins differ extensively, except for those of HP1 and HP2, and are believed to be involved in interactions with the host proteins required for DNA replication. Both phage P2 and phage 186 require the host proteins DnaB, DnaE, DnaG and Rep (18, 60). Genetic evidence indicates a direct interaction between the Rep helicase and a glutamic acid residue at position 155 of the phage 186 A protein (152). Phage 186 also requires DnaC, the DnaB helicase loader, in contrast to phage P2 where the B protein has been shown to interact with DnaB and is thus believed to act as a helicase loader (113). However, since P2 is able to form a functional minichromosome in the absence of protein B, the B protein requirement is only valid for phage replication. Phage 186 is unable to replicate at restrictive temperature in *dnaA*-ts mutants of *E. coli*, but this has been shown to be an indirect requirement since phage 186 will replicate in integratively suppressed *dnaA*-defective *E. coli* strains (136, 152).

The phage P2 A protein has another intriguing property, namely that it works preferentially in *cis*. This was initially detected as a lack of complementation of *A* amber mutants by simultaneous wild-type infections in phages S13, ΦX174, G4, and P2 (41, 83, 135, 143, 144). That this is a function of the *A* gene alone has been shown using phage P2 "A" minichromosomes which contain only the *A* gene and an antibiotic resistance marker (113a). Cells containing a wild-type minichromosome cannot be transformed with a minichromosome containing an *A* amber mutation under su^- conditions, while this occurs with a high frequency under su^+ conditions. Furthermore, under in vitro conditions, using coupled transcription–translation, the *cis* preference of the phage P2 A protein for its own gene is maintained (113a). The mechanism behind the preferred *cis* activity for phage ΦX174 has been suggested to be immediate membrane entrapment upon synthesis (41, 148). But this cannot be the case for phage P2 since the same preference occurs under in vitro conditions. Purified phage P2 A protein is unable to cleave covalently closed circular double-stranded DNA or linear double-stranded DNA. Thus, the coupled transcription–translation per se or some component present in the S30 extract is required for the *cis*-activity. Possibly the A protein acts on its own DNA target before it is fully translated.

Activation of Late Transcription

P2 late genes are transcribed from four late promoters, designated P_P, P_O, P_V and P_F, while those of phage 186 are transcribed from the three promoters P_{12}, P_V, and P_J (corresponding to the phage P2 promoters P_P, P_O, and P_F, respectively). As can be seen in figure 25-4, the promoters have poor similarities to the consensus −10 and −35 regions of the *E. coli* σ^{70}-dependent promoters. Instead they have a region with a partial dyad symmetry centered approximately 55 bp downstream the transcriptional initiation site. This dyad symmetry has been shown by deletion analysis and base substitutions to be essential for promoter activity (46, 147).

Initiation of late gene transcription requires active DNA replication, and a transcriptional activator, designated Ogr in P2 and B in 186 (13, 24–26). Ogr and B have been shown to be members of a family of small transcriptional activators containing a C_2C_2 zinc finger motif that are present in some members of the P2 family of phages: NucC from a cryptic prophage in *Serratia marcescens* (64), Pag from the *Salmonella* phage PSP3 (66), Ogr from the *Pseudomonas* phage ΦCTX (103), and ORF13 from *Vibrio cholerae* phage K139 (72) (figure 25-5). However, no similar protein has been found in neither phage HP1 nor phage HP2 of *Haemophilus*, indicating that these phages activate late gene transcription by another mechanism (35). The late genes of P2, and most likely phage186, can also be directly activated by the δ proteins of satellite phages P4 and ΦR73 (66). The δ proteins are members of the same family of zinc containing proteins as Ogr and B. The δ protein of phage P4, however, is twice the size of the P2 Ogr protein and contains two zinc fingers. It appears to be a covalent dimer of Ogr (51, 68) (figure 25-5). The binding of the transcriptional activators phage PSP3 Pag, phage P4 δ and phage ΦR73 δ to the phage P2 late promoters, and the B protein binding to the phage 186 P_V promoter, has been analyzed by DNaseI footprinting and the results confirm the binding of the activators to the dyad symmetry around position −55 (66, 67, 117).

The phage P2 Ogr, phage 186 B, and phage ΦR73 δ proteins have been purified. The assumption that the conserved cysteine residues are involved in forming a complex with Zn(II) is supported by ^{65}Zn blotting and atomic absorption spectroscopy analysis (69, 79, 117). Moreover, site-directed mutagenesis of the conserved cysteines of phage P2 Ogr

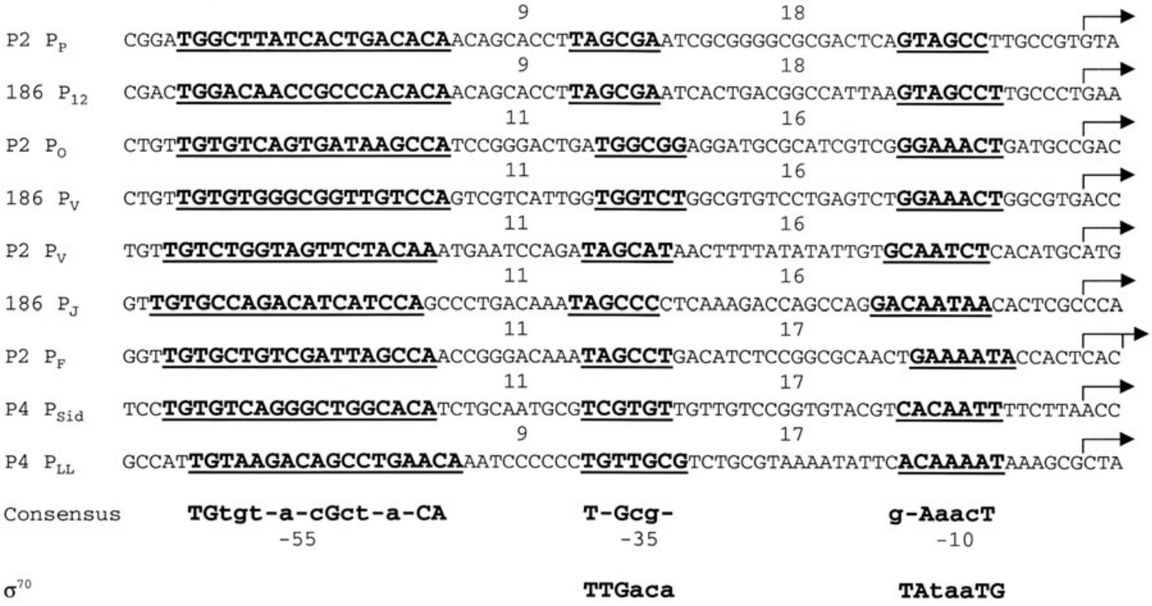

P2 P_P CGGATGGCTTATCACTGACACAACAGCACCTTAGCGAATCGCGGGGCGCGACTCAGTAGCCTTGCCGTGTA
186 P_{12} CGACTGGACAACCGCCCACACAACAGCACCTTAGCGAATCACTGACGGCCATTAAGTAGCCTTGCCCTGAA
P2 P_O CTGTTGTGTCAGTGATAAGCCATCCGGGACTGATGGCGGAGGATGCGCATCGTCGGGAAACTGATGCCGAC
186 P_V CTGTTGTGTGGGCGGTTGTCCAGTCGTCATTGGTGGTCTGGCGTGTCCTGAGTCTGGAAACTGGCGTGACC
P2 P_V TGTTGTCTGGTAGTTCTACAAATGAATCCAGATAGCATAACTTTTATATATTGTGCAATCTCACATGCATG
186 P_J GTTGTGCCAGACATCATCCAGCCCTGACAAATAGCCCCTCAAAGACCAGCCAGGACAATAACACTCGCCCA
P2 P_F GGTTGTGCTGTCGATTAGCCAACCGGGACAAATAGCCTGACATCTCCGGCGCAACTGAAAATACCACTCAC
P4 P_{Sid} TCCTGTGTCAGGGCTGGCACATCTGCAATGCGTCGTGTTGTTGTCCGGTGTACGTCACAATTTTCTTAACC
P4 P_{LL} GCCATTGTAAGACAGCCTGAACAAATCCCCCCTGTTGCGTCTGCGTAAAATATTCACAAAATAAAGCGCTA
9
18
11
16
17
Consensus TGtgt-a-cGct-a-CA T-Gcg- g-AaacT
-55 -35 -10
σ^{70} TTGaca TAtaaTG

Figure 25-4 Comparison of the promoter regions of the late operons in phages P2, 186, and P4. The transcriptional start sites are indicated by the arrows. The common sequences are indicated in bold, and their consensus sequences are indicated below the promoter sequences and compared with conserved −10 and −35 regions of the σ^{70} promoter.

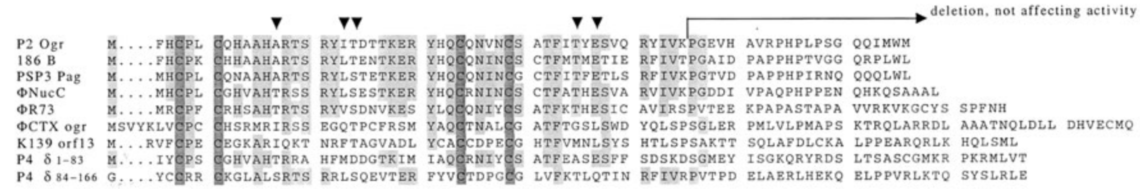

deletion, not affecting activity
P2 Ogr M....FHCPL CQHAAHARTS RYITDTTKER YHQCQNVNCS ATFITYESVQ RYIVKPGEVH AVRPHPLPSG QQIMWM
186 B M....FHCPK CHHAAHARTS RYLTENTKER YHQCQNINCS CTFMTMETIE RFIVTPGAID PAPPHPTVGG QRPLWL
PSP3 Pag M....MHCPL CQNAAHARTS RYLSTETKER YHQCQNINCG CTFITFETLS RFIVKPGTVD PAPPHPIRNQ QQQLWL
ΦNucC M....MHCPL CGHVAHTRSS RYLSESTKER YHQCRNINCS CTFATHESVA RVIVKPGDDI VPAQPHPPEN QHKQSAAAL
ΦR73 M....MRCPF CRHSAHTRTS RYVSDNVKES YLQCQNIYCS ATFKTHESIC AVIRSPVTEE KPAPASTAPA VVRKVKGCYS SPFNH
ΦCTX ogr MSVYKLVCPC CHSRMRIRSS EGQTPCFRSM YAQCTNALCG ATFTGSLSWD YQLSPSGLER PMLVLPMAPS KTRQLARRDL AAATNQLDLL DHVECMQ
K139 orf13 M...RVFCPE CEGKARIQKT NRFTAGVADL YCACCDPECG HTFVMNLSYS HTLSPSAKTT SQLAFDLCKA LPPEARQRLK HQLSML
P4 δ 1-83 M....IYCPS CGHVAHTRRA HFMDDGTKIM IAQCRNIYCS ATFEASESFF SDSKDSGMEY ISGKQRYRDS LTSASCGMKR PKRMLVT
P4 δ 84-166 G....YCCRR CKGLALSRTS RRLSQEVTER FYVCTDPGCG LVFKTLQTIN RFIVRPVTPD ELAERLHEKQ ELPPVRLKTQ SYSLRLE

Figure 25-5 Alignment of the transcriptional activators for the late promoters of P2 and some P2-like phages. Alignment was performed using the CLUSTAL X program (59). The conserved cysteins are indicated by dark-gray shading, amino acids common to at least five proteins in light gray. Amino acids believed to be involved in interactions with the RNA polymerase α are indicated with arrowheads.

and phage ΦR73 δ proteins also shows that the cysteines are required for biological activity (43, 44, 69). In fact, even replacing the C_2C_2 motif with a C_2H_2 zinc binding motif in ΦR73 δ leads to a complete loss of zinc binding and biological activity (69).

The Ogr-like transcriptional activators are believed to interact with the host RNA polymerase α, since P2 Ogr is inactive in *E. coli* strains with specific mutations in the *rpoA* gene (3, 42, 140). By isolating mutants of phages P2, P4 and ΦR73 able to grow on such *E. coli rpoA* mutants the amino acids equivalent to P2 Ogr residues number 13, 19, 20, 42, and 44 have been implicated to interact with the α subunit of RNA polymerase (52, 69, 74). These residues are not conserved in all Ogr-like proteins (figure 25-5), but they are located in regions that are conserved among the members. The N-terminal two thirds of the protein is well conserved among the Ogr-like proteins, but not the C-terminus. This fits with the finding that the 21 C-terminal amino acid residues are not required for phage P2 Ogr protein activity (44).

The P2 *ogr* and 186 *B* genes are transcribed from two different promoters. A normal *E. coli* σ^{70} promoter, which should be active early after infection, precedes both genes. Late in infection they are transcribed instead as part of one of the late tail operons (14, 71). In phage 186, the expression of *B* is controlled by the cI protein, while the P2 Ogr protein seems to be under indirect immunity control (14, 26). Furthermore, in phage 186 it has been shown that activation of the late promoter requires only the B protein and that DNA replication is not necessary (25). Thus, phage P2's requirement for DNA replication for late-promoter activation might be explained by an Ogr-protein dose effect.

Lysis

All double-stranded phages studied so far use a holin–endolysin system for host cell lysis (154, chapter 10). The endolysins can be grouped according to their muralytic activities. Transglycosidases and lysozymes attack the glycosidic bond, and the amidases and endopeptidases attack the amide and peptide bonds (155). In most cases, the holin and the endolysin genes are localized in a lysis cassette together with genes encoding accessory lysis proteins. Bacteriophage P2 and the P2-like phages studied so far have two essential lysis genes. First is the endolysin gene designated *K*, *P*, *lys* or ORF28 in phages P2, 186, HP1 and HP2, and K139, respectively, which is orthologous to the λ *R* gene encoding a transglycosylase. The second is the holin gene designated *Y*, ORF24 or *hol* in phages P2, 186, and HP1 and HP2 (154, 160) (table 25-1). The holin of phage HP1, Hol, differs from Y and ORF24 of phages P2 and 186. A holin for phage K139 has not yet been identified. This might not be surprising considering the extensive diversity of holin genes. So far, more than 35 unrelated orthologous gene families have been identified (150; chapter 10).

Phage P2 has two ancillary lysis genes, *lysA* and *lysB*, which are functionally homologous to the *Rz* and *Rz1* proteins, although *lysA* and *lysB* are two separate but adjacent genes while *Rz1* is embedded within *Rz* in a different reading frame (154, 160). Mutants in *lysA* cause slightly delayed lysis and might encode an antiholin, while mutants in *lysB* show accelerated lysis (160). Genes homologous to *lysB* are also found in phage 186 and phage ΦCTX, where they seem to play a role in the correct timing of lysis.

Lysogenization

All P2-like phages studied so far integrate into the host chromosome upon lysogenization. The integration is mediated by a phage-encoded integrase that promotes recombination between a phage attachment site (*attP*) and a bacterial attachment site (*attB*), generating host–phage junctions, designated *attL* and *attR*. This site-specific recombination leads to no loss or gain of nucleotides. Integration also requires the integration host factor, IHF, which acts as an architectural protein by bending the DNA. The reverse event, excision, requires an additional phage-encoded protein. Thus, the P2-like phages use the same mechanism for integration as the well-studied λ site-specific recombination system, but the phage proteins and their DNA binding sites differ (78).

Chromosomal Insertion Sites

The preferred integration sites of phage phages P2, 186, and WΦ in the *E. coli* genome (*attB*) differ in sequence and in location (table 25-2), which is noteworthy since these integration sites are recognized by phage-encoded integrases that are presumed to have a common ancestry. This suggests that there has been selection in favor of different integration sites, allowing several P2-like prophages to occupy the same host. A strongly preferred attachment site for phage P2 has been found only in *E. coli* C strains. The corresponding sites on the maps of *E. coli* K-12 and B strains are occupied by defective P2-like prophages. As a consequence, phage P2 integrates at many secondary sites, which show up to 37% mismatches within the core sequence (6, 138).

The integrases of the P2-like phages analyzed so far are all members of the integrase family of recombinases, which recognize inverted repeated sequences in the host chromosome. The cleavage-joining reaction is believed to occur in two steps. First one strand in *attP* and one in *attB* are cleaved, exchanged, and joined to each other. The second cleavage-joining takes place after a short branch migration has occurred. The sequence recognized by the integrase, the core sequence, is therefore denoted BOB′ in the host, where B and B′ are the sequences recognized by the integrase and O is the intervening region where branch migration takes place. Among the P2-like phages, the crossover points have been determined only for phage HP1, which gives a branch migration of 7 bp (50) (table 25-2). Interestingly, the core sequence of phage ΦCTX contains only a direct repeat, and in the case of phage P2 the inverted repeat constituting the B and B′ sequences have a poor identity. By mutational analysis the right half, B′, has been shown to be the primary recognition sequence (C. Frumerie, A. Yu and E. Haggård-Ljungquist, unpublished data).

Many phages integrate into tRNA genes. To avoid disruption of the gene, the phage contains a long identity region so that part of the tRNA is duplicated upon integration and a complete copy of the gene is maintained. It has been suggested that the primordial *attB* sites were tRNA genes (23). As can be seen in table 25-2, phages P2 and WΦ do not integrate into tRNA genes. Instead, they have short identity regions located in spacer regions between *E. coli* genes, and phage K139 integrates between the *flaC* gene and the gene encoding the flagellin core protein, A, in *V. cholerae*.

Structure of the Phage Attachment Sites

The site-specific recombination systems of phages P2, 186, WΦ, and HP1 are very similar to the λ system, where the integrase and the host integration factor (IHF) assemble at the phage attachment site, *attP*, forming the recombinogenic intasome that unites into a synaptic complex with the host *attB* site (78). Like λ integrase, the integrases of phages P2, 186, WΦ, and HP1 also have two different DNA binding epitopes, one recognizing the core DNA sequence and the other the arm sequences (36, 87, 158). The reverse event, excision, requires an additional phage-encoded protein that in λ is designated Xis. In phages P2, 186, WΦ, and HP1, the equivalent function is provided by the repressor of the lysogenic promoter: the Cox protein in phage P2, in phage WΦ and in phage HP1, and the Apl protein in phage 186 (30, 36, 37, 87, 159). However, as can be seen in figure-25.6, there are differences in the size of the *attP* regions and in the number and location of the binding sites of the proteins required for recombination in the respective phage. Phage HP1 has the largest *attP* site, spanning about 400 bp,

Table 25-2 Host Integration Sites of P2-Related Phages

Phage	Host	Integration Site (kb from *ori*)	Gene(s) or ORF(s)	AttB Sequence[a]	Reference/ Accession No.
P2	*E. coli* K-12	2 165.2	Between *yegQ* and *b2083*	AAAAAATAAGCCCGTGTAAGGGAGATT	156
P2	*E. coli* O157:E7	2 842.6	Between *Ec2889* and *Ec2890*	AAAAAATAAGCCCGTGTAAGGGAGATT	54
P2	*E. coli* O157:E7 EDL933	2 912.8	Between *Z3250 (yegQ)* and *Z3251*	AAAAAATAAGCCCGTGTAAGGGAGATT	113
WΦ	*E. coli* K-12	4 103.9	Between *cpxR* and *pfkA*	GACACCATCCCTGTCTTCCCCCACATGCTGTGGGGGTTTTTTTTATC	86
186	*E. coli* K-12	2 783.8 and 3 213.3	Within *ileY* and *ileX* tRNA$^{\mathrm{ile}}$	TGCTGGACTTGAACCAGCGACCAAGCGATTATGAGT	U32222, NC_001317
HP1	*Haemophilus influenzae*	91.8 and 139.6	Within tRNA$^{\mathrm{leu}}$	AGGGA ↓ TTTTAAA ↓ TCCCTT	52, 53
ϕCTX	*Pseudomonas aeruginosa*	2 947.6	Within tRNA$^{\mathrm{ser}}$	ATATGGCGGAGGCGG TGAGATTCGAACTC	55
K139	*Vibrio cholerae*	2 334.4 (chromosome I)	Between *FlaC* and Flagellin core protein A	CAGAAAAGGGGCTTTTCTTTTTTC[b]	104

[a] The cleavage sites in HP1 are indicated by arrows. Inverted repeats are underlined with a single line, direct repeats are underlined with a double line.

[b] Not verified experimentally.

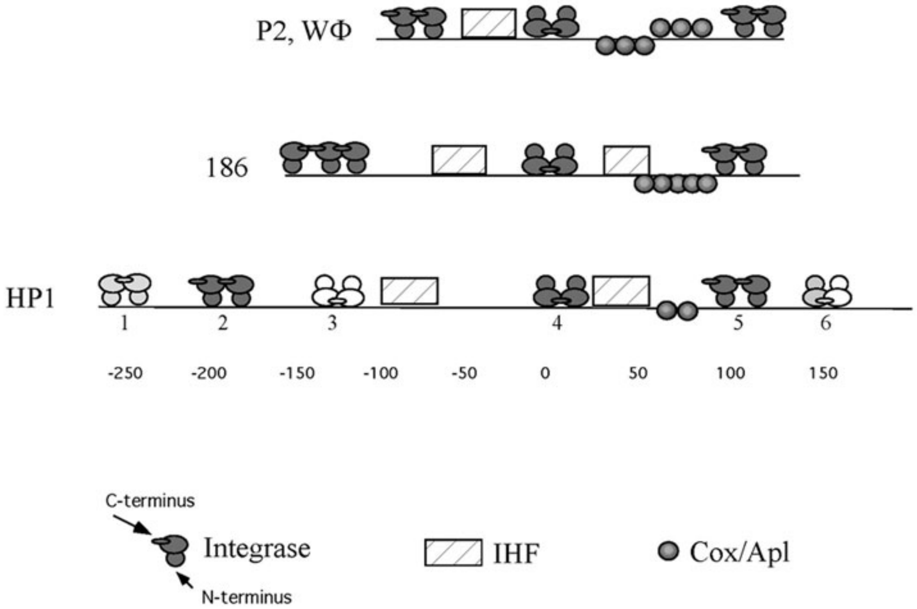


Figure 25-6 A schematic drawing of the *attP* regions of phage P2 and some P2-like phages. Shown are the relative binding sites of proteins required for integration and excision. In phage HP1, the dark gray indicate integrase binding sites required for recombination, those in light gray have stimulatory function, and white binding sites are nonessential. See text for more details.

whereas the equivalent site in phage P2 is about half that size. Compared with phage P2, phage HP1 has additional integrase binding sites (38, 49, 158). The C-terminal DNA binding epitope, which recognizes the core sequence, has two additional binding sites on either side of the core. The left site is dispensable and the right site is required for excision but not for integration. The phage HP1 integrase also has an extra arm-binding site, about 250 bp to the left of the core, which is stimulatory. It should be noted that this site has an inverted repeat, in contrast to the other two arm-binding sites which have direct repeats. Both phages HP1 and 186 have two IHF binding sites, one on each side of the core, while phage P2 only has one site, which is to the left of the core (28, 61, 158).

All three phages have the Cox/Apl binding site located to the right of the core sequence, but they differ in both binding-site number and orientations. Phage HP1 has only two Cox-binding sites, located as a directly repeated sequence of 10 bp, spaced by 1 bp, and one Cox tetramer binds to each repeat (37). The phage P2 Cox protein has six repeated sequences of 9 bp, oriented so that they can form a long inverted repeat with three repeats in each arm (159). In phage 186, the Apl protein recognizes five repeated sequences of 6 bp, all in the same orientation and located so that they will be exposed on the same face of the DNA helix (30). Phage WΦ has the integrase and IHF binding sites located at similar positions compared with phage P2, but the core sequences differ while the arm sequences are identical (87) (table 25-2). The location of the phage WΦ Cox protein binding sites has not been determined.

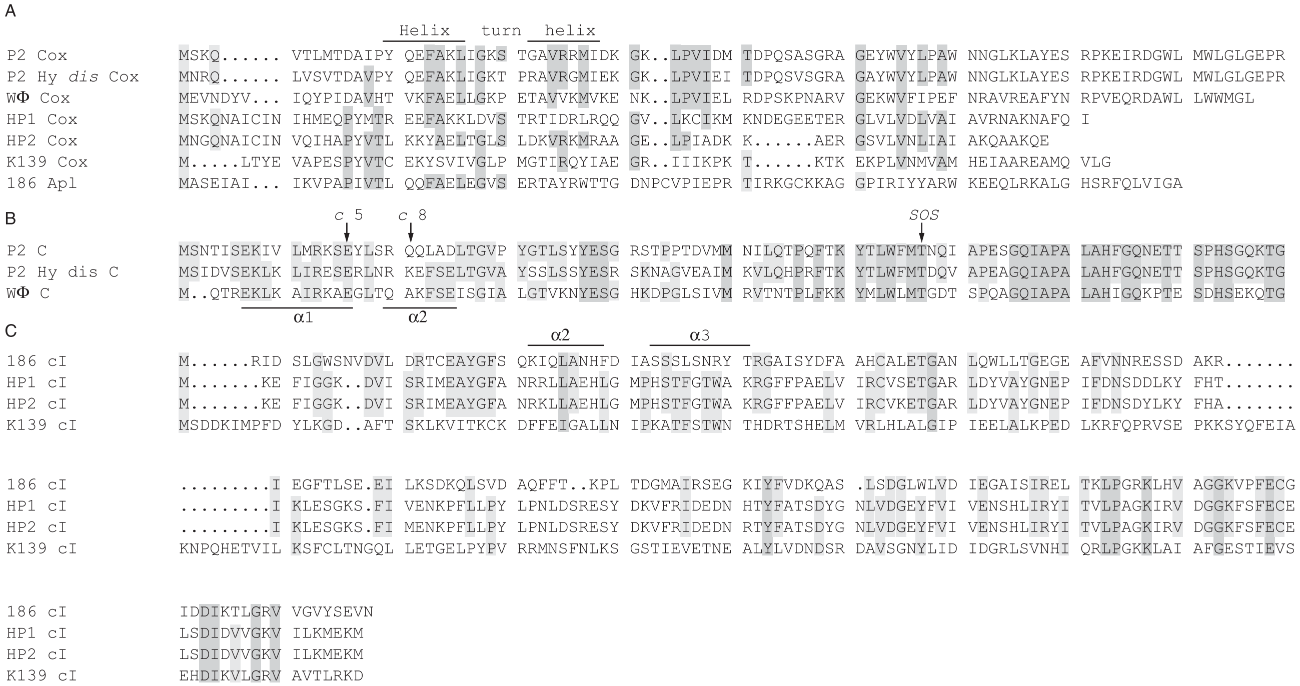
Phage Proteins Involved in Site-Specific Recombination

The alignment of 105 site-specific recombinases of the integrase protein family has revealed two conserved boxes and three conserved patches of charged amino acids (110) (figure 25-7). Box I contains the invariant Arg residue, and Box II contains the conserved His-Xxx-Xxx-Arg motif, and the active site tyrosine, which in phage ΦCTX is shifted 3 amino acids compared with the others. The phage λ integrase, the prototype for genetic and biochemical studies of the integrase family of recombinases, has at least two domains: a small N-terminal domain that has a high affinity for the arm-type sites in *attP*, and a large C-terminal domain that contains both the low affinity core-binding site and the catalytic site. The DNA bending protein, IHF, functions by binding to the region between the arm and the core, producing a U-turn that brings the integrase binding sites into close proximity. This allows the integrase bound to the high-affinity arm sites to bind to the low-affinity core site (73, 122). The structures of the catalytic domains of the phage λ and phage HP1 integrases have been determined, and they contain a globular domain composed of a bundle of α-helices with a three- or four-stranded antiparallel β-sheet on the outside (21, 76). In the monomer structure of phage HP1 integrase, the 17-residue-long C-terminal tail extends away from the globular domain, and the dimers observed in the crystal are formed by interactions of the C-terminal tails that orient the active-site clefts antiparallel to each other (21). Both phage HP1 and phage λ integrases

```
P2 Int            M....AIKKL DDGR...YEV DIRPTGRNGK RIRRKFDKKS EAVAFEKYTL
WΦInt             M....SIKKL DDGR...YEV DIRPRGRDGK RIRRKFERKA EAVAFERYTI
K139 Int          M....SVRNL KDGSKKPWLC ECYPQGREGK RVRKRFATKG EATAYENFIM
186 Int           M....TVRKN PAGG...WIC ELYPNGAKGK RIRKKFATKG EALAF.....
HP1 Int           M....AVRKD TKNGK..WLA EVYVN...GN ASRKWFLTKG DALRFYNQAK
HP2 Int           M....AVRKD TKNGK..WLA EVYVN...GK RLRKWFLTKG DALRFYNQAK
ΦCTX Int          MADGVEVRGK RIRIYFRYQG ELCRESIPGD ATPENIANAE RLAGIINYEI

P2 Int            YNHH...NKE WLSK.PTDKR RLSELTQIWW DLKGKHEEHG KSN...LGKI
WΦInt             AYAS...QKE WAGQ.RADRR TLSELLDIWW KYHGQNHEHG TKE...FNYL
K139 Int          REVD...DKP WMGS.KPDNR RLSELLETWW QVHGHTIKSG KVV...YRKT
186 Int           EQYT..VQNP WQEE.KEDRR TLKELVDSWY SAHGITLKDG LKR...QLAM
HP1 Int           EQTTSAVDSV QVLE.SSDLP ALSFYVQEWF DLHGKTLSDG KAR...LAKL
HP2 Int           EQTTSAVDSV QVLE.SSDLP ALSFYVQEWF DLHGKTLSDG EAR...LAKL
ΦCTX Int          KQGVFSYSRH FPDSPRVKSN TLGHYIDLWL DIKRNQIAAS GFRGYTSRVE

P2 Int            EIFTKITNDP CAFQITKSLI SQYCATRR.. .......SQG IKPSSINRDL
WΦInt             LKTISGIGDI PVSRMSKRAL MDYRSMRI.. .......RDG ISAATINRDM
K139 Int          ALTIKELGDP IASTFTSKQY LAFRASRVS. ..HFNKENKS LSPTYQNFQL
186 Int           HHAFECMGEP LARDFDAQMF SRYREKRLKG EYARSNRVKE VSPRTLNLEL
HP1 Int           KNLCSNLGDP PANEFNAKIF ADYRKRRLDG EFS.VNKNNP PKEATVNREH
HP2 Int           KNLCANLGDP PANEFNAEIF ADYRKRRLDG EFS.VNKNNP PKEATVNREH
ΦCTX Int          THIRPRWGDS QADSIDHLDI QDWVQNTLMP .........K LHNKTVREIV

P2 Int            TCISGMFTAL IEAELFFGEH PIRGTKRLKE EKPETGYLTQ EEISLLLAAL
WΦInt             YRLSGMFTKL IQLDEFSGQH PIHGLPPLAE ANPEMTFLEK AEIEKLLNVL
K139 Int          NLLSGMFSRL IKYKQWNLPN PLDDIEPIKV NQRALAYLDK ADIQLPFQRL
186 Int           AYFRAVFNEL NRLGEWKGEN PLKNMRPFRT EEMEMTWLTH DQISQLLGEC
HP1 Int           AYLRAVFNEL KSLRKWTTEN PLDGVRLFKE RETELAFLYE RDIYRLLAEC
HP2 Int           AYLRAVFNEL KSLRKWTAEN PLDGVRLFKE RETELAFLYE RDIYRLLAEC
ΦCTX Int          SNLRQIFR.L YRTRNRSAHD PTDGIVITLP DADDPDPFTR EEIDLILGTE
                                                            Patch I
P2 Int            DGDN...... .....KKIAL LCLSTGARWG EAARLKAE.. .........N
WΦInt             AGDD...... .....LLVAL LCLSTGGRWT EVATLKPA.. .........Q
K139 Int          GGFESDGRSV SIPEIVLIAK ICLATGARIS EALSLERS.. .........Q
186 Int           NRHDH..... ..PDLETVVR ICLATGARWS EAESLRKS.. .........Q
HP1 Int           DNSRN..... ..PDLGLIVR ICLATGARWS EAETLTQS.. .........Q
HP2 Int           DNSRN..... ..PDLGLIVR ICLATGARWS EAETLTQS.. .........Q
ΦCTX Int          TARIG..... ....ELNLAE FMIWSGPRVS EAIALAWEDV DLDTGTVVFR
                                                  Box I
P2 Int            IIHNRVTFVK TKTNK..... .......... .......PRT VPISEAVAKM
WΦInt             ITNCRVTFLK TKNGK..... .......... .......KRT VPISEELEKK
K139 Int          ISEFKLTFVE TKGKR..... .......... .......IRS VPISENLYKE
186 Int           LAKYKITYTN TKGRK..... .......... .......NRT VPISKELYES
HP1 Int           VMPYKITFTN TKSKK..... .......... .......NRT VPISKELFDM
HP2 Int           VMPYKITFTN TKSKK..... .......... .......NRT VPISKELFDM
ΦCTX Int          RARVRSQYKV TKTRRSTRKV QLLAPALRAL QQQAKLTRRL PPVQIEVIDR
                                         Patch II
P2 Int            IADNKRG... .FLFPDADYP RFRRTMKAIK PDLPMGQAT. ........HA
WΦInt             VKEEASA... .KLF.KVDYE KFCGILRRVK PDIPPNQAT. ........HI
K139 Int          IMLASSSS.. TKIFSTTYGS AHRYIKKALP DYVPEGQAT. ........HV
186 Int           LPDDKKG... .RLFSDCYGA .FRSALERTG IELPAGQLT. ........HV
HP1 Int           LPK.KRG... .RLFNDAYES .FENAVLRAE IELPKGQLT. ........HV
HP2 Int           LPK.KRG... .RLFNDAYES .FENAVLRAE IELPKGQLT. ........HV
ΦCTX Int          DNRTRKPQRV RFVFHNSASG AAYSTSDTLR NGWWHGHLRN AGVRSRGPNO
                          Patch III
P2 Int            LRHSFATHFM INGGSIITLQ RILGHT.RIE QTMVYAHFAP EYLQDAISLN
WΦInt             LRHTFASHFM MNGGNIIALQ QILGHA.SIQ QTMAYAHLAP DYLQNAVALN
K139 Int          LRHTFATHFM MNRGDILILQ RVLGHQ.KIE QTMAYAHFSP DHLIQAVQLN
186 Int           LRHTFASHFM MNGGNILVLQ EILGHS.TIE MTMRYAHFAP DHLEDAVKLN
HP1 Int           LRHTFASHFM MNGGNILVLK EILGHS.TIE MTMRYAHFAP SHLESAVKFN
HP2 Int           LRHTFASHFM MNGGNILVLK EILGHS.TIE MTMRYAHFAP SHLESAVKFN
ΦCTX Int          CRHTFASQML SSG..IATPE WIADQMGHTS TAMIFKHYAK WISKDGPDIV
                                Box II
P2 Int            PLRGGTEAES VHTVSTVE
WΦInt             PLKGGVTL
K139 Int          PLEN
186 Int           PLVHITNSK
HP1 Int           PLSNPAQ
HP2 Int           PLSNPAQ
ΦCTX Int          GLLNQALKLS
```

Figure 25-7 Alignment of integrase proteins of phage P2 and some P2-like phages. The amino acids were aligned using the CLUSTAL X program (59). Amino acids present in at least four proteins are indicated by gray shading. The conserved boxes and patches among the integrase family of proteins are underlined (110).

have been shown to form dimers and oligomers in solution (49). However, in the case of λ integrase, it seems to be the N-terminus that is involved in protein–protein interactions (63). Using genetic and biochemical analyses, phage P2 integrase has also been shown to form dimers and in this case the C-terminus as well as central parts seem to be involved in protein–protein interactions (C. Frumerie, J. M. Eriksson, M. Dugast, and E. Haggård-Ljungquist,



```
A
                                        Helix   turn  helix
P2 Cox          MSKQ...... VTLMTDAIPY QEFAKLIGKS TGAVRRMIDK GK..LPVIDM TDPQSASGRA GEYWVYLPAW NNGLKLAYES RPKEIRDGWL MWLGLGEPR
P2 Hy dis Cox   MNRQ...... LVSVTDAVPY QEFAKLIGKT PRAVRGMIEK GK..LPVIEI TDPQSVSGRA GAYWVYLPAW NNGLKLAYES RPKEIRDGWL MWLGLGEPR
WΦ Cox          MEVNDYV... IQYPIDAVHT VKFAELLGKP ETAVVKMVKE NK..LPVIEL RDPSKPNARV GEKWVFIPEF NRAVREAFYN RPVEQRDAWL LWWMGL
HP1 Cox         MSKQNAICIN IHMEQPYMTR EEFAKKLDVS TRTIDRLRQQ GV..LKCIKM KNDEGEETER GLVLVDLVAI AVRNAKNAFQ I
HP2 Cox         MNGQNAICIN VQIHAPYVTL KKYAELTGLS LDKVRKMRAA GE..LPIADK K......AER GSVLVNLIAI AKQAAKQE
K139 Cox        M.....LTYE VAPESPYVTC EKYSVIVGLP MGTIRQYIAE GR..IIIKPK T......KTK EKPLVNMVAM HEIAAREAMQ VLG
186 Apl         MASEIAI... IKVPAPIVTL QQFAELEGVS ERTAYRWTTG DNPCVPIEPR TIRKGCKKAG GPIRIYYARW KEEQLRKALG HSRFQLVIGA

B                                c 5   c 8                                                  SOS
                                ↓     ↓                                                     ↓
P2 C            MSNTISEKIV LMRKSEYLSR QQLADLTGVP YGTLSYYESG RSTPPTDVMM NILQTPQFTK YTLWFMTNQI APESGQIAPA LAHFGQNETT SPHSGQKTG
P2 Hy dis C     MSIDVSEKLK LIRESERLNR KEFSELTGVA YSSLSSYESR SKNAGVEAIM KVLQHPRFTK YTLWFMTDQV APEAGQIAPA LAHFGQNETT SPHSGQKTG
WΦ C            M..QTREKLK AIRKAEGLTQ AKFSEISGIA LGTVKNYESG HKDPGLSIVM RVTNTPLFKK YMLWLMTGDT SPQAGQIAPA LAHIGQKPTE SDHSEKQTG
                             α1          α2
C                                                     α2          α3
186 cI          M......RID SLGWSNVDVL DRTCEAYGFS QKIQLANHFD IASSSLSNRY TRGAISYDFA AHCALETGAN LQWLLTGEGE AFVNNRESSD AKR.......
HP1 cI          M.......KE FIGGK..DVI SRIMEAYGFA NRRLLAEHLG MPHSTFGTWA KRGFFPAELV IRCVSETGAR LDYVAYGNEP IFDNSDDLKY FHT.......
HP2 cI          M.......KE FIGGK..DVI SRIMEAYGFA NRKLLAEHLG MPHSTFGTWA KRGFFPAELV IRCVKETGAR LDYVAYGNEP IFDNSDYLKY FHA.......
K139 cI         MSDDKIMPFD YLKGD..AFT SKLKVITKCK DFFEIGALLN IPKATFSTWN THDRTSHELM VRLHLALGIP IEELALKPED LKRFQPRVSE PKKSYQFEIA


186 cI          .........I EGFTLSE.EI LKSDKQLSVD AQFFT..KPL TDGMAIRSEG KIYFVDKQAS .LSDGLWLVD IEGAISIREL TKLPGRKLHV AGGKVPFECG
HP1 cI          .........I KLESGKS.FI VENKPFLLPY LPNLDSRESY DKVFRIDEDN HTYFATSDYG NLVDGEYFVI VENSHLIRYI TVLPAGKIRV DGGKFSFECE
HP2 cI          .........I KLESGKS.FI MENKPFLLPY LPNLDSRESY DKVFRIDEDN RTYFATSDYG NLVDGEYFVI VENSHLIRYI TVLPAGKIRV DGGKFSFECE
K139 cI         KNPQHETVIL KSFCLTNGQL LETGELPYPV RRMNSFNLKS GSTIEVETNE ALYLVDNDSR DAVSGNYLID IDGRLSVNHI QRLPGKKLAI AFGESTIEVS


186 cI          IDDIKTLGRV VGVYSEVN
HP1 cI          LSDIDVVGKV ILKMEKM
HP2 cI          LSDIDVVGKV ILKMEKM
K139 cI         EHDIKVLGRV AVTLRKD
```

Figure 25-8 Alignments of amino acid sequences. Alignments were performed using the CLUSTAL X program (59). A: Cox proteins. The location of the helix–turn–helix structure believed to be involved in DNA binding is indicated. The location of a P2 *cox* mutant, defective in dimerization, is indicated by a star. Amino acids present in at least four proteins are indicated by gray shading. B: Immunity C proteins of P2 and P2-like phages. The two α-helices presumed to contain the DNA binding motif are indicated. The mutations discussed in the text are indicated above the sequence. Amino acids common to all three proteins are shaded in dark gray, those present in two are shaded in light gray. C: Immunity cI proteins of 186 and the 186-like proteins. Amino acids common to at least four proteins are shaded in dark gray, those present in three are shaded in light gray. The location of the two α-helices, involved in DNA binding, are indicated.

unpublished data). When the integrases of the P2-like phages are aligned they show some homology outside the conserved boxes and patches, even though very few amino acids are conserved in all proteins. It should be noted that the integrases of HP1 and HP2 are almost identical (figure 25-7).

As noted above, in the P2-like phages the repressors of the lysogenic promoters also have an architectural role during excisive recombination. However, in the case of P2 and HP1, the Cox proteins are also involved in regulating the direction of recombination since they are not only required for excisive recombination but also inhibit the integrative recombination (37, 159). Binding of the Cox protein to *attP* in phage HP1 has been shown to prevent binding of integrase to the right arm site, which explains its inhibitory effect on integrative recombination. A similar analysis has not been performed with the phage P2 Cox protein, but P2 Cox binding has been shown to bend the DNA target about 72° (J. M. Eriksson and E. Haggård-Ljungquist, unpublished data). This bending might affect binding of Int protein to the P′ arm since the footprints of the Cox and Int proteins are slightly overlapping (158). The DNA-binding domains of the Cox/Apl proteins are believed to be a helix–turn–helix motif located at the N-terminal end of the respective protein (29, 37, 129). This region is also the most conserved part of the proteins (figure 25-8). The native forms of the Cox proteins of phages HP1 and P2 are tetramers that can self-associate to octamers, while protein Apl of phage 186 is a monomer in solution (34, 37, 134). It has been shown with phage P2 Cox protein that the protein–protein interacting interface is located in the C-terminal region and that oligomerization is necessary for biological activity (34).

The fact that the integrase and the *cox* genes are located in two different, mutually exclusive transcriptional units—the lysogenic and the early transcript, respectively—poses a problem since they both are needed for excision of the prophage after derepression. Thus, the spontaneous lysogen induction is low, but measurable (7). However, derepression of a P2 lysogen with a temperature-sensitive repressor is abortive: less than 1% of the bacteria will produce phage (10). This has been shown to be due to insufficient production of integrase from the prophage (93). The level of expression of P2 integrase is affected at several steps. First, the Pc promoter is rather weak and under negative control by the Cox protein. Secondly, there is a partial transcriptional terminator, located between gene *C* and *int*, that allows only 30% read-through. Thirdly, the final transcriptional terminator is located downstream of *attP*. This means that the *C-int* transcript lacks a terminator signal after integration and will continue into an untranslated region in the host chromosome, which might make the transcript accessible for RNase attacks. Finally, the integrase functions as a translational repressor by binding to its own transcript covering the ribosomal binding site (156).

A Model for Intasome Formation

The HP1 site-specific recombination system has been extensively analyzed. By mutating the different DNA binding sites of the proteins involved in recombination, a tentative model for intasome formation has been generated (38). The basic steps for integrative recombination are: (i) The C-terminal ends of two integrase molecules will bind to the core sequence (site 4 in figure 25-6), (ii) the N-terminus of each core binding integrase molecule binds to one of the repeats in the respective arm (site 2 and site 5) in a process that is stimulated by binding of IHF, (iii) two other integrase molecules will bind to the free arm repeats with their N-terminal ends at sites 2 and 5, and (iv) their C-terminus will bind to the *attB* core sequence allowing recombination.

Control of Lytic Versus Lysogenic Growth

The developmental switch of temperate phages must be set so that the phage after infection enters either the lytic cycle or the lysogenic cycle. The two pathways must be mutually exclusive. There are three common characteristics of the transcriptional switches of the P2-like phages: (i) two face-to-face promoters that control the lytic or lysogenic functions with partially overlapping transcripts; (ii) two repressors, the immunity repressor and the Cox/Apl repressors, which recognize different operators and control each other's promoter; and (iii) the Cox/Apl repressors also act as excisionases. However, the P2-like phages have two types of immunity repressors, based on size, composition, and structure of the target. The establishment and maintenance of immunity also differs among members of the group, since some have a cII-like function and some respond to the host SOS function. A schematic drawing of the switches of phages P2 and 186 are shown in figure 25-9.

The Transcriptional Switch

Even if the transcriptional switches of the P2-like phages have similar arrangements, several differences are evident. Enlarged representations of the promoter-operator regions are shown in figure 25-10. Phage HP1 is unique since it has two early promoters, P_{R1} and P_{R2}, but as P_{R1} is relatively weak it has been suggested to play only a minor role in vivo (39). The two face-to-face promoters that control the switches give transcripts that overlap for different lengths for the different phages. The phage P2 transcripts have the shortest overlap of about 35 bp (A. Ahlgren-Berg and E. Haggård-Ljungquist, unpublished data). Depending on which promoter is used in phage HP1, the transcripts overlap with 44 or 72 bp (39). In phage 186 the overlap is about 60 bp; and other phages in the group have overlaps of 72–75 bp (29, 87, 105, 120).

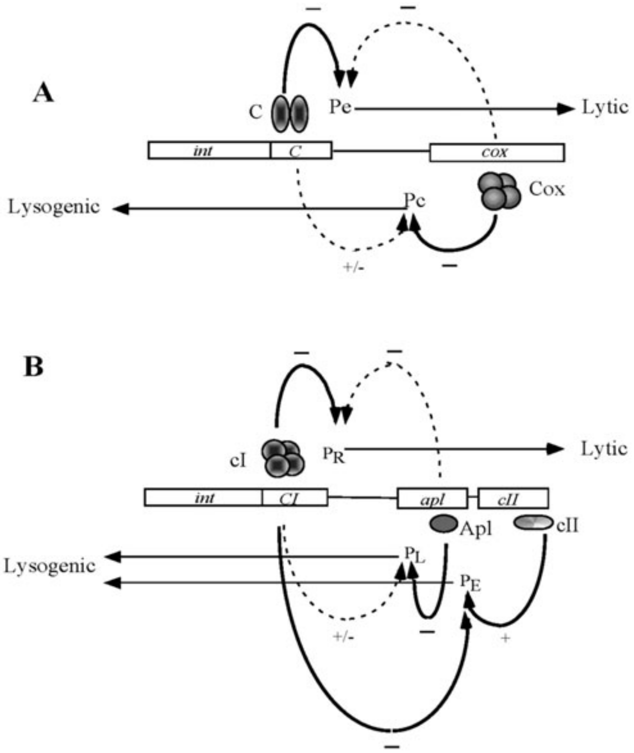


Figure 25-9 A schematic drawing of the transcriptional switches. A: Phage P2. B: Phage 186. See text for more details.

The early promoters are much stronger than the lysogenic promoters for all phages except HP1, where the two promoters are of similar strength (29, 39, 87, 120, 130). In all phages the strength of the lysogenic promoter is significantly reduced (10- to 20-fold) by the presence of the lytic promoter. In phages P2 and 186 the interference of the early transcript in the activity of the lysogenic transcript can be abolished by addition of immunity repressor, but this has not been found in HP1 (29, 39, 87, 120, 130).

As can be seen in figure 25-8, the immunity proteins of the P2-like phages can be divided into two types based on size and sequence similarity: the P2-like and the 186-like proteins. The three C proteins of the P2 type are small, homologous proteins. The native form of phage P2 C protein is a dimer, and the C-terminal part is believed to be involved in dimerization (89, 121). The C-terminal ends are very similar, and the C proteins of phages P2 and WΦ can form heterodimers that are able to repress a hybrid operator containing both phage WΦ O1 and phage P2 O2 operators, which neither protein can accomplish alone (P. Peltola and E. Haggård-Ljungquist, unpublished data). Thus, two operators and dimerization of the C protein are required for

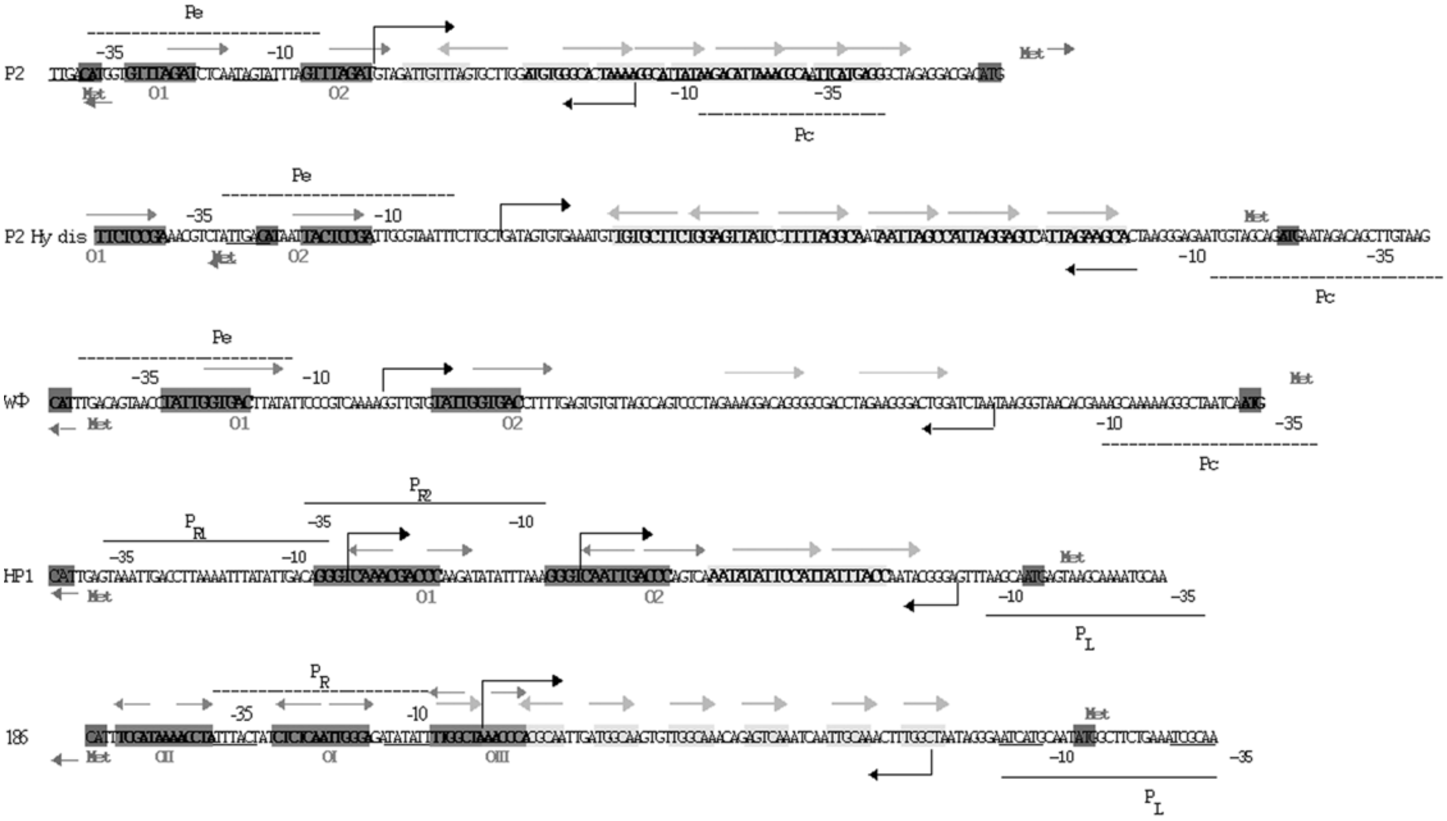

Pe
-35
-10
P2
O1
O2
Met
-10
-35
Pc
Pe
-35
-10
P2 Hy dis
O1
Met
O2
Met
-10
-35
Pc
Pe
-35
-10
WΦ
Met
O1
O2
Met
-10
-35
Pc
P_{R2}
-35
-10
P_{R1}
-35
-10
HP1
Met
O1
O2
Met
-10
-35
P_L
P_R
-35
-10
186
Met
OII
OI
OIII
Met
-10
-35
P_L

Figure 25-10 The nucleotide sequences of the promoter-operator regions of phage P2 and some phage P2-like phages. The −10 and −35 regions are underlined, and the start of the transcripts are indicated with bent arrows. The operators for the immunity repressors are indicated by light gray shading, and those of the Cox/Apl repressors in dark gray shading. The arrows above the shaded operators indicate the orientation of the respective recognition sequence for each protein. The start codons of the genes are indicated.

biological activity. The N-terminal ends of the C proteins contain two predicted α-helices believed to be involved in DNA binding. The phage P2 *c5* and the phage P2 Hy *dis vir14* mutations, which are located in the first α-helix, change the conserved glutamic acid residue at position 15 to a valine and a lysine, respectively (120). Another mutation in phage P2, *c8*, is located in the second α-helix that is less conserved, as is expected if it constitutes the DNA recognition helix. The *c8* mutation has been shown not to affect dimerization, which fits with the hypothesis that the N-terminus constitutes the DNA binding motif, and that the dimerization interface is located in the C-terminus (121).

The immunity repressors of the P2 type recognize direct repeats of different lengths, spanning either the −10 or the −35 region of the respective promoter. Also the distances between the operators differ. In phages P2 and WΦ they are located on the same face of the DNA helix, since the distance from center to center of the repeats is 22 and 33 bp respectively (87, 94). In phage P2 Hy *dis*, however, the distance is 26 bp, and they will therefore not be on the same face of the DNA (120).

The cI immunity repressors of the phage 186 type are about twice the size of the C repressors of the phage P2 type, and the two subgroups show no sequence similarity (37, 70, 105) (figure 25-8). A recent search for proteins homologous to phage 186 cI repressor has identified several other phages/prophages, in different hosts, containing related proteins (133), but they have not yet been analyzed for other properties in common with the P2-like phages and are thus not discussed here. The cI repressors of phages 186 and HP1 have been shown to recognize inverted repeat sequences. The cI protein of phage 186 has been purified and the native form was shown to consist of dimers that self-associate via tetramers to octamers (134). The protein has further been shown to consist of two domains, where the N-terminal domain contains the DNA-binding motif and the C-terminal domain contains the oligomerization surface (27, 133). Besides controlling the early lytic promoter P_R, the phage 186 cI repressor controls the expression of the late promoter activator protein B and the expression of cII. In addition, a third binding site within the *cI* gene has been identified (27). Phage 186 cI repressor has three binding sites within P_R: one to the left of the −35 region, one between the −35 region and the −10 region, and a third to the right of the −10 region. Each operator is repeated at a distance of two turns of the helix (27, 77) (figure 25-10). It should be noted that the central operator of phage 186 (OI) has a consensus sequence that differs from that of the adjecent operators (OII and OIII), and by mutational analysis it has been shown that the α-helices 2 and 3 at the N-terminus, which constitute a weak helix–turn–helix motif, recognize both operator sequences (27, 133). Phage HP1 has two predicted operators spanning the −10 region of promoter P_{R2} (39). The CI proteins of phages 186, HP1, HP2, and K139 show similarities in size and amino acid sequence, but the number of conserved amino acids are few, and most are located in the C-terminal part of the proteins (35, 70, 105) (figure 25-8). The cI proteins of HP1 and HP2, however, are almost identical.

The repressors that control the lysogenic promoter of the respective switch are the multifunctional Cox/Apl proteins, which also act as architectural proteins during excisive recombination. The Cox/Apl proteins recognize a different DNA sequence compared with the immunity proteins. For phage P2, the Cox recognition sequences overlap with the Pc promoter. For all other P2-like phages, they are located downstream of the −10 region (figure 25-10). Thus, with phage P2 the Cox protein may block RNA polymerase binding, but in the other P2-like phages the Cox/Apl proteins may affect steps later than RNA polymerase binding such as open-complex formation or promoter clearance. The Cox proteins, with the exception of phage HP1 Cox, also autoregulate their own expression (37, 87, 129), but this does not appear to be of biological significance since Cox mutants show normal lytic growth. The P2 Cox protein has been shown to bend its DNA target upon binding, and if this is also true for the other Cox/Apl proteins then it can explain the effect on both promoters (J. M. Eriksson and E. Haggård-Ljungquist, unpublished data).

The Cox proteins of phages P2, P2 Hy *dis*, and WΦ are more similar to each other than to the Cox/Apl proteins of phages 186, HP1, HP2, and K139 (figure 25-8). The most conserved region is the helix–turn–helix motif in the N-terminal part of the proteins, believed to be involved in DNA binding. It is noteworthy that the HP1 and HP2 Cox proteins are similar only at the N- and C-termini, while the interiors of the two proteins show no similarity. This is in contrast to the cI, Int, and A proteins, which are almost identical in phages HP1 and HP2.

Satellite phage P4, which needs all P2 late functions for lytic growth (see chapter 26), needs a signal to recognize the presence of a helper phage. When phage P2 infects a P4-lysogenic strain, the P2 Cox protein is used as a transcriptional activator that turns on the phage P4 Pll promoter that controls P4 DNA replication (128). The P4 Pll promoter is normally activated by the P4 δ gene product (139). The phage P2 Cox protein has no similarity to phage P4 δ protein, but acts by mimicking the way δ bypasses the normal P4 immunity, that is by binding to an extended region upstream of the Pll promoter. This region contains up to six Cox recognition sequences that are all positioned in the same direction (128). The Cox protein of phage P2 Hy *dis* is also able to activate the phage P4 Pll promoter, in contrast to the phage WΦ Cox protein that is unable to do so (87, 120).

Establishment of Lysogeny

Phage 186 has been found to have an additional gene (*cII*) that controls the establishment but not maintenance of lysogeny. A homologous gene is present at the same location in phage K139 (72), but the function of the cII protein has been studied only in phage 186. The phage 186 *cII* gene is located in the early operon, downstream of gene *apl*, and encodes a transcriptional activator that acts on the P_E promoter located between genes *apl* and *cII* (77, 106, 107) (figure 25-9). The P_E promoter, which is negatively controlled by the cI protein, has in the presence of the cII protein about the same strength as the P_R promoter and it increases transcription of *cI* at least 55 times (106, 107). The P_E promoter, as opposed to the P_L promoter, is not affected by the Apl protein (107). The cII protein recognizes an inverted repeat sequence, separated by two helical turns of the DNA helix, that is located upstream of the P_E promoter. The cII protein has a predicted helix–turn–helix motif at the N-terminal end, believed to be the DNA binding site.

The control of establishment of lysogeny of phage 186 is very similar to that of phage λ, since they both have a cII protein that functions as an activator of a promoter located upstream of the weak P_L promoter that enhances transcription of the cI protein. Even though the cII proteins have a similar function in these phages, they have no detectable amino acid sequence homology and the transcriptional switches controlling the maintenance of these phages are very different (118).

The fact that phages 186 and P2 have similar lysogenization frequencies, even though P2 lacks a cII protein, is difficult to explain. With phage P2 the lytic promoter is at least 100 times stronger than the lysogenic promoter, while in phage 186 the cII-activated P_E promoter makes the opposing promoters almost equal in strength. Thus, other factors involved in the control of these molecular switches will most likely be revealed in the future.

Induction

Bacteriophage P2 has been the prototype for the noninducible class of temperate phages: indeed, its immunity repressor protein lacks the sequence present in proteins that are induced to self-cleavage by the activated RecA protein. But phage 186 is UV-inducible, even though its immunity repressor is insensitive to the activated RecA (71). Instead, the induction of the 186 prophage is dependent on a phage-encoded protein designated Tum. The *tum* gene is located at the left end of the genome and it is transcribed from the P_{95} promoter together with the open reading frame ORF97. The P_{95} promoter is controlled by the LexA protein, which is sensitive to the activated RecA protein (19). The Tum protein has been purified and shown to be an antirepressor that binds to the cI protein, preventing it from binding to the operator (132).

Phage P2 and its satellite phage, P4, have the capacity to mutually derepress each other. As described above, derepression of a P4 lysogen by phage P2 is mediated by the P2 Cox protein that acts as a transcriptional activator of the late P4 promoter, Pll. When phage P4 infects a P2 lysogen, it needs to derepress prophage P2 in order to gain access to the P2 late genes. This is accomplished by the P4 *ε* gene, which promotes derepression of prophage P2 leading to in situ phage-P2 DNA replication and activation of late genes (45). The E protein has been shown to act as an antirepressor since it binds to the C protein, but not to DNA, leading to a shift from the lysogenic to the lytic mode (88). Like the C protein, the E protein forms homodimers but not multimers. However, in the presence of both proteins, multimeric complexes of E and C are formed that interfere with the binding of C to its operator (89). The phage P2 *sos* mutant grows normally, but in the prophage state it is no longer derepressed by the P4 E protein. P2 *sos* will, however, support P4 growth in a mixed infection (5). The *sos* mutation is located within the coding part of the *C* gene where it changes threonine at position 67 to isoleucine, and the E protein has been shown to be unable to bind to the mutated C protein (121). The C proteins of phages WΦ and P2 Hy *dis* also respond to the E protein, which fits with the fact that the threonine at position 67 is located in a conserved region of the C proteins (figure 25-8). The cI proteins of phages 186, HP1, HP2, and K139 lack this protein domain, and are not derepressed by P4 E. This explains why prophage 186 is not derepressed upon a P4 infection (131).

Evolution of P2-Like Phages

Phage genomes in general can be characterized as mosaics containing both genes similar to other phage genes, or host genes, and genes showing no similarity to any known gene (15, 58). P2-like phages are no exception; they are similar in many aspects but all contain unique genes, some with unknown function. A phage is taxonomically classified as P2-like if it shares some, but necessarily not all, characters with phage P2 (1). Many phages have been found to comply with this criterion but there are only six complete genomes available (table 25-1).

Phylogenetic Relationships

A genome comparison of these six fully sequenced P2-like phages reveals that it is only an integrase gene and nine late genes (corresponding to genes *Q, P, O, N, M, L, S, H,* and *T* in P2) that are both genetically similar and present in all six genomes. Separate phylogenetic analysis of the amino acid sequences of the proteins encoded by the nine late genes all result in the same tree, which implies that they share the same evolutionary history (figure 25-11). The closer two phages are positioned in the phylogenetic tree based on all nine genes, the more similar the genes they share in the rest of the genome, for example phage P2 compared with phage 186 (table 25-1). There is no indication of major recombination events between these nine homologous genes for any pair of phages, which suggests that it is

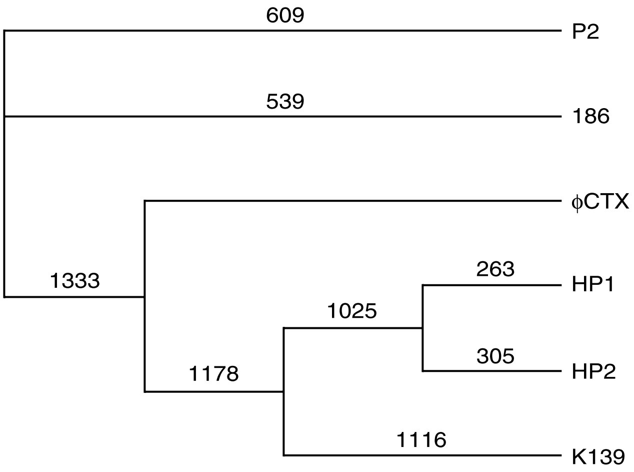


Figure 25-11 Phylogeny of P2 and various P2-like phages. Unrooted phylogenetic tree showing the relationship between inferred amino acid sequences from nine late genes, homologous to phage P2 genes *Q, P, O, N, M, L, S, H,* and *T*. The tree was constructed with maximum-parsimony criteria and was the shortest tree found using the exhaustive search procedure. Numbers above branches indicate branch length (total 7549; the next shortest tree had a total length of 7732). The total number of characters was 4281 and there were 1913 parsimoniously informative characters. The tree was constructed with the program PAUP, version 4.0b8a (141).

likely these genes are inherited clonally (A. S. Nilsson, unpublished data).

The phylogenetic relationship of other genes, particularly the early regulatory genes, is often ambiguous and their evolutionary history cannot be resolved. Phylogenetic analyses of the amino acid sequences of the six integrase proteins (figure 25-8) produce very weak trees in which phages HP1/HP2 and 186 harbor the most similar *int* genes. Some informative characters also speak for a close relationship between the integrases of phages K139 and 186, as well as between phages K139 and P2. The relationship of the *int* gene of phage ΦCTX to the other integrases is impossible to establish. Phylogenetic analyses of the amino acid sequence of the A protein show that phages P2 and 186 are closely related in the central conserved part of the protein and that phage K139 is related to phages HP1 and HP2, but these relationships shift more than once in the last third of the protein (A. S. Nilsson, unpublished data) (figure 25-8). Although the number of informative characters is rather low, recombination is a more plausible explanation for these homoplasies than convergent evolution or recurring neutral mutations.

Homologous and Non-Homologous Recombination

Homologous recombination has been shown to be a more important mechanism than mutation for nucleotide change in phage P2, which is not surprising since about 30% of the *E. coli* strains contain P2-like prophages and exchange of genetic information is known to occur between host genomes (40). A study of five late genes from 18 closely related P2-like phage isolates demonstrated that homologous recombination is extensive and happens at random breakpoints (108). The amount of genetic variation was low in these genes. Pairwise comparisons of the nucleotide sequence of the most differentiated gene all showed over 96% identity. Since there was disproportionately more variation in synonymous rather than nonsynonymous third-codon positions, it was suggested that these late genes are subjected to rather strong stabilizing selection.

The fact that homologous recombination is detectable between slightly differentiated late genes but not between more different late genes, could mean that recombination does not occur in the latter because different capsids belong to evolutionarily separate clones with different life histories, for example different host preference, and hardly ever meet. Another explanation is that recombination does occur but that the capsid proteins are too different, leading to assembly of nonfunctional capsids which are quickly eliminated by selection. Indeed, the amino acid sequences of the capsid genes, for instance those of phages P2 and ΦCTX, are very different and few are informative. Hence, a third explanation is that minor recombination events may pass unnoticed, which leads to the wrong conclusion (A. S. Nilsson, unpublished data).

It is also known that recombination can occur between two phage P2 mutants in a mixed infection, although at an extremely low frequency and in a *recA*-independent manner. The frequency of ordinary, *recA*-dependent homologous recombination, either between two prophages or by non-replicating P2 phages infecting a lysogen, is much higher and seems to be somewhat more frequent around the phage P2 origin of replication (9).

Phage evolution is not only dependent on homologous recombination between related phages. Many similar genes, or parts of genes, are frequently found in otherwise unrelated phages. A comparison of tail fiber genes from P2, P1, Mu, λ, K3, and T2 demonstrates regions of similarity, which indicates a high level of non-homologous recombination between unrelated phages (47). For example, the protein assumed to be involved in tail fiber assembly, G, is virtually identical (93%) in phages P2 and Mu (protein U), but the two phages share only around 43% of their protein-G amino acids with ORF45, the corresponding protein in phage 186, which is supposed to be the closest relative of P2.

Morons

A phage genome often contains nonessential genes that are missing in close relatives. Such genes often differ in AT/GC ratio when compared with the rest of the genome and they are seldom transcriptionally coupled to other genes but instead carry their own promoters and intrinsic transcriptional terminators. These are traits indicating that these genes are recent additions to the genome capable of being autonomously transcribed (57). These phage genes are sometime referred to as morons (65), and there are at least three genes—*Z/fun*, *old*—and *tin*, that belong to this class in the P2 genome. Morons that are transcribed and translated from an integrated prophage (that is, lysogenic conversion) may affect the fitness of the host. This seems to apply to the three P2 genes since they all affect *E. coli* and make the lysogen refractory to T5, λ, and T-even phages, respectively (22, 101, 102).

Many of the P2-like phages in *E. coli* isolates of the ECOR collection have different inserts at the locus corresponding to P2 *Z/fun*. Amplification of the region between the *G* and F_I genes, using only one set of primers, gave 12 fragments of which nine were of different length (1.2–3.6 kb). Sequencing of some of these fragments has shown that the genetic variation is high. There seem to be only two pairs of prophages having the same insert, and many contain easily identified open reading frames with a high A-T content. Some of these morons have been shown to carry their own promoters and their expression has been confirmed in coupled transcription–translation assays, but the functions of these genes are not known. The morons are inserted at the same position, in the spacer region between genes *G* and F_I, creating sharp boundaries between similar and unique sequences in an alignment. Thus the mechanism behind the insertions appears to be site-specific (109). Phages 186 and PSP3 contain nothing but a short noncoding sequence between the genes equivalent to *G* and F_I, but phage WΦ has a 1.4 kb insert that is 61% AT. Phage WΦ is very similar to phage P2 in many respects but it has a putative methylase gene at the same position as the *old* and *tin* morons in phage P2 (G. E. Christie, personal communication).

Note Added in Proof

Relevant publications since this review was written:

Phage 186

Callen, B. P., K. E. Shearwin, and J. B. Egan. 2004. Transcriptional interference between convergent promoters caused by elongation over the promoter. Mol. Cell. 14:647–656.

Dodd, I. B., and J. B. Egan. 2002. Action at a distance in CI repressor regulation of the bacteriophage 186 genetic switch. Mol. Microbiol. 45:697–710.

Phage P2

Christie, G. E., D. L. Anders, V. McAlister, T. S. Goodwin, B. Julien, and R. Calendar. 2003. Identification of upstream sequences essential for activation of a bacteriophage P2 late promoter. J. Bacteriol. 185:4609–4614.

Christie, G. E., L. M. Temple, B. A., Bartlett, and T. S. Goodwin. 2002. Programmed translational frameshift in the bacteriophage P2 *FETUD* tail gene operon. J. Bacteriol. 184:6522–6531.

Frumerie, C., J. M. Eriksson, D. Dugast, and E. Haggård-Ljungquist. 2005. Dimerization of bacteriophage P2 integrase is not required for binding to its DNA target but for it biological activity. Gene 344:221–231.

Frumerie, C., L. Sylwan, A. Ahlgren-Berg, and E. Haggård-Ljungquist. 2005. Cooperative interactions between bacteriophage P2 integrase and its accessory factors IHF and Cox. Virology 332:284–294.

Markov, D., G. E. Christie, B. Sauer, R. Calendar, T. Park, R. Young, and K. Severinov. 2004. P2 growth restriction on an *rpoC* mutant is suppressed by alleles of the *Rzl* homolog *lysC*. J. Bacteriol. 186:4628–4637.

Evolution

Nilsson, A. S., J. L. Karlsson, and E. Haggård-Ljungquist. 2004. Site-specific recombination links the evolution of P2-like coliphages and pathogenic enterobacteria. Mol. Biol. Evol. 21:1–13.

Kita, K., H. Kawakami, and H. Tanaka. 2003. Evidence for horizontal transfer of the *Eco*T38I restriction-modification gene to chromosomal DNA by the P2 phage and diversity of defective P2 prophages in *Escherichia coli* TH38 strains. J. Bacteriol. 185:2296–2305.

Williams, B. J., M. Golomb, T. Phillips, J. Brownlee, M.V. Olson, and A. L. Smith. 2002. Bacteriophage HP2 of *Haemophilus influenzae*. J. Bacteriol. 184:6893–6905.

Acknowledgments

We wish to thank all our friends and colleagues for contributing unpublished results, for critical reading of the manuscript, and for helpful suggestions. This work is based on knowledge available March 2002.

References

1. Ackermann, H. W. 1999. Tailed bacteriophages: The order *Caudovirales*. Adv. Virus Res. 51:135–201.
2. Agarwal, M., M. Arthur, R. D. Arbeit, and R. Goldstein. 1990. Regulation of icosahedral virion capsid size by the in vivo activity of a cloned gene product. Proc. Natl. Acad. Sci. USA 87:2428–2432.
3. Ayers, D. J., M. G. Sunshine, E.W. Six, and G. E. Christie. 1994. Mutations affecting two adjacent amino acid residues in the alpha subunit of RNA polymerase block transcriptional activation by the P2 Ogr protein. J. Bacteriol. 176:7430–7438.

4. Baas, P. D., F. Heidekamp, A. D. M. Van Mansfeld, H. S. Jansz, S. A. Langeveld, G. A. Van der Martel, G. H. Veeneman, and J. H. Von Boom. 1981. Essential features of the origin of bacteriophage ΦX174 RF DNA replication: a recognition and a key sequence for ΦX gene A protein, pp. 195–209. *In* D. S. Ray (ed.) The Initiation of DNA Replication. Academic Press, New York.
5. Barclay, S. L., and W. F. Dove. 1978. Mutations of bacteriophage P2 which prevent activation of P2 late genes by satellite phage P4. Virology 91:321–335.
6. Barreiro, V., and E. Haggård-Ljungquist. 1992. Attachment sites for bacteriophage P2 on the *Escherichia coli* chromosome: DNA sequences, localization on the physical map, and detection of a P2-like remnant in *E. coli* K-12 derivatives. J. Bacteriol. 174:4086–4093.
7. Bertani, G. 1962. Multiple lysogeny from single infection. Virology 18:131–139.
8. Bertani, G. 1951. Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. J. Bacteriol. 62:293–300.
9. Bertani, G., and G. Dehò. 2001. Bacteriophage P2: recombination in the superinfection preprophage state and under replication control by phage P4. Mol. Genet. Genomics 266:406–416.
10. Bertani, L. E. 1968. Abortive induction of bacteriophage P2. Virology 36:87–103.
11. Bertani, L. E., and G. Bertani. 1971. Genetics of P2 and related phages. Adv. Genet. 16:199–237.
12. Bertani, L. E., and E. W. Six. 1988. The P2-like phages and their parasite, P4, pp. 73–143. *In* R. Calendar (ed.) The Bacteriophages, vol. 2. Plenum Publishing, New York.
13. Birkeland, N. K., and B. H. Lindqvist. 1986. Coliphage P2 late control gene *ogr*. DNA sequence and product identification. J. Mol. Biol. 188:487–490.
14. Birkeland, N. K., B. H. Lindqvist, and G. E. Christie. 1991. Control of bacteriophage P2 gene expression: analysis of transcription of the *ogr* gene. J. Bacteriol. 173:6927–6934.
15. Botstein, D. 1980. A theory of modular evolution for bacteriophages. Ann. N.Y. Acad. Sci. 354:484–491.
16. Bowden, D. W., and R. Calendar. 1979. Maturation of bacteriophage P2 DNA in vitro: a complex, site-specific system for DNA cleavage. J. Mol. Biol. 129:1–18.
17. Bowden, D. W., and P. Modrich. 1985. In vitro maturation of circular bacteriophage P2 DNA. Purification of *ter* components and characterization of the reaction. J. Biol. Chem. 260:6999–7007.
18. Bowden, D. W., R. S. Twersky, and R. Calendar. 1975. *Escherichia coli* deoxyribonucleic acid synthesis mutants: their effect upon bacteriophage P2 and satellite bacteriophage P4 deoxyribonucleic acid synthesis. J. Bacteriol. 124:167–175.
19. Brumby, A. M., I. Lamont, I. B. Dodd, and J. B. Egan. 1996. Defining the SOS operon of coliphage 186. Virology 219:105–114.
20. Bullas, L. R., A. R. Mostaghimi, J. J. Arensdorf, P. T. Rajadas, and A. J. Zuccarelli. 1991. *Salmonella* phage PSP3, another member of the P2-like phage group. Virology 185:918–921.
21. Burgess Hickman, A., S. Waninger, J. J. Scocca, and F. Dyda. 1997. Molecular organization in site-specific recombination: the catalytic domain of bacteriophage HP1 integrase at 2.7 Å resolution. Cell 89:227–237.
22. Calendar, R., S. Yu, H. Myung, V. Barreiro, R. Odegrip, K. Carlson, L. Davenport, G. Mosig, G. Christie, and E. Haggård-Ljungquist. 1998. The lysogenic conversion genes of coliphage P2 have unusually high AT content, pp. 241–252. *In* C. I. Kado (ed.) Horizontal Gene Transfer. Chapman and Hall, London.
23. Campbell, A. M. 1992. Chromosomal insertion sites for phages and plasmids. J. Bacteriol. 174:7495–7499.
24. Christie, G. E., E. Haggård-Ljungquist, R. Feiwell, and R. Calendar. 1986. Regulation of bacteriophage late-gene expression: the *ogr* gene. Proc. Natl. Acad. Sci. USA 83: 3238–3242.
25. Dibbens, J. A., and J. B. Egan. 1992. Control of gene expression in the temperate coliphage 186. IX. *B* is the sole phage function needed to activate transcription of the phage late genes. Mol. Microbiol. 6:2629–2642.
26. Dibbens, J. A., S. L. Gregory, and J. B. Egan. 1992. Control of gene expression in the temperate coliphage 186. X. The *cI* repressor directly represses transcription of the late control gene *B*. Mol. Microbiol. 6:2643–2650.
27. Dodd, I. B., and J. B. Egan. 1996. DNA binding by the coliphage 186 repressor protein CI. J. Biol. Chem. 271: 11532–11540.
28. Dodd, I. B., and J. B. Egan. 1999. P2, 186 and related phages (*Myoviridae*), pp. 1087–1094. *In* A. Granoff and R. G. Webster (eds.) Encyclopedia of Virology, 2nd edn, vol. 1. Academic Press, San Diego.
29. Dodd, I. B., B. Kalionis, and J. B. Egan. 1990. Control of gene expression in the temperate coliphage 186. VIII. Control of lysis and lysogeny by a transcriptional switch involving face-to-face promoters. J. Mol. Biol. 214: 27–37.
30. Dodd, I. B., M. R. Reed, and J. B. Egan. 1993. The Cro-like ApI repressor of coliphage 186 is required for prophage excision and binds near the phage attachment site. Mol. Microbiol. 10:1139–1150.
31. Dokland, T. 1999. Scaffolding proteins and their role in viral assembly. Cell. Mol. Life Sci. 56:580–603.
32. Dokland, T., B. H. Lindqvist, and S. D. Fuller. 1992. Image reconstruction from cryo-electron micrographs reveals the morphopoietic mechanism in the P2–P4 bacteriophage system. EMBO J. 11:839–846.
33. Doran, T. J., S. M. Loh, N. Firth, and R. A. Skurray. 1994. Molecular analysis of the F-plasmid *traVR* region: *TraV* encodes a lipoprotein. J. Bacteriol. 176:4182–4186.
34. Eriksson, J. M., and E. Haggård-Ljungquist. 2000. The multifunctional bacteriophage P2 Cox protein requires oligomerization for biological activity. J. Bacteriol. 182: 6714–6723.
35. Esposito, D., W. P. Fitzmaurice, R. C. Benjamin, S. D. Goodman, A. S. Waldman, and J. J. Scocca. 1996. The complete nucleotide sequence of bacteriophage HP1 DNA. Nucleic Acids Res. 24:2360–2368.
36. Esposito, D., and J. J. Scocca. 1994. Identification of an HP1 phage protein required for site-specific excision. Mol. Microbiol. 13:685–695.
37. Esposito, D., and J. J. Scocca. 1997. Purification and characterization of HP1 Cox and definition of its role in

controlling the direction of site-specific recombination. J. Biol. Chem. 272:8660–8670.

38. Esposito, D., J. S. Thrower, and J. J. Scocca. 2001. Protein and DNA requirements of the bacteriophage HP1 recombination system: a model for intasome formation. Nucleic Acids Res. 29:3955–3964.

39. Esposito, D., J. C. E. Wilson, and J. J. Scocca. 1997. Reciprocal regulation of the early promoter region of bacteriophage HP1 by the Cox and CI proteins. Virology 234:267–276.

40. Feil, E. J., E. C. Holmes, D. E. Bessen, M.-S. Chan, N. P. J. Day, M. C. Enright, R. Goldstein, D. W. Hood, A. Kalia, C. E. Moore, J. Zhou, and B. G. Spratt. 2001. Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. Proc. Natl. Acad. Sci. USA 98:182–187.

41. Francke, B., and D. S. Ray. 1972. *Cis*-limited action of the gene-A product of bacteriophage ΦX174 and the essential bacterial site. Proc. Natl. Acad. Sci. USA 69:475–479.

42. Fujiki, H., P. Palm, W. Zillig, R. Calendar, and M. Sunshine. 1976. Identification of a mutation within the structural gene for the α subunit of DNA-dependent RNA polymerase of *E. coli*. Mol. Gen. Genet. 145:19–22.

43. Gebhardt, K. 1994. Structure and function analysis of the bacteriophage P2 Ogr protein. PhD thesis University of Oslo, Oslo, Norway.

44. Gebhardt, K., R. A. King, G. E. Christie, and B. H. Lindqvist. 1993. Mutational analysis of bacteriophage P2 Ogr protein: truncation of the carboxy terminus. J. Bacteriol. 175:7724–7726.

45. Geisselsoder, J., P. Youderian, G. Dehò, M. Chidambaran, R. Goldstein, and E. Ljungquist. 1981. Mutants of satellite virus P4 that cannot derepress their bacteriophage P2 helper. J. Mol. Biol. 148:1–19.

46. Grambow, N. J., N. K. Birkeland, D. L. Anders, and G. E. Christie. 1990. Deletion analysis of a bacteriophage P2 late promoter. Gene 95:9–15.

47. Haggård-Ljungquist, E., C. Halling, and R. Calendar. 1992. DNA sequences of the tail fiber genes of bacteriophage P2: evidence for horizontal transfer of tail fiber genes among unrelated bacteriophages. J. Bacteriol. 174:1462–1477.

48. Haggård-Ljungquist, E., E. Jacobsen, S. Rishovd, E. W. Six, Ø. Nilssen, M. G. Sunshine, B. H. Lindqvist, K.-J. Kim, V. Barreiro, E. V. Koonin, and R. Calendar. 1995. Bacteriophage P2: genes involved in baseplate assembly. Virology 213:109–121.

49. Hakimi, J. M., and J. J. Scocca. 1994. Binding sites for bacteriophage HP1 integrase on its DNA substrates. J. Biol. Chem. 269:21340–21345.

50. Hakimi, J. M., and J. J. Scocca. 1996. Purification and characterization of the integrase from the *Haemophilus influenzae* bacteriophage HP1: identification of a four-stranded intermediate and the order of strand exchange. Mol. Microbiol. 21:147–158.

51. Halling, C., and R. Calendar. 1990. Bacteriophage P2 *ogr* and P4 δ genes act independently and are essential for P4 multiplication. J. Bacteriol. 172:3549–3558.

52. Halling, C., M. G. Sunshine, K. B. Lane, E. W. Six, and R. Calendar. 1990. A mutation of the transactivation gene of satellite bacteriophage P4 that suppresses the *rpo*A109 mutation of *Escherichia coli*. J. Bacteriol. 172: 3541–3548.

53. Hauser, M. A., and J. J. Scocca. 1992. Site-specific integration of the *Haemophilus influenzae* bacteriophage HP1: identification of the points of recombinational strand exchange and the limits of the host attachment site. J. Biol. Chem. 267:6859–6864.

54. Hauser, M. A., and J. J. Scocca. 1992. Site-specific integration of the *Haemophilus influenzae* bacteriophage HP1: location of the boundaries of the phage attachment site. J. Bacteriol. 174:6674–6677.

55. Hayashi, T., K. Makino, M. Ohnishi, K. Kurakowa, K. Ishii, K. Yokoyama, C.-G. Han, E. Ohtsubo, K. Nakayama, T. Murata, M. Tanaka, T. Tobe, T. IIda, H. Takami, T. Honda, C. Sasakawa, N. Ogasawara, T. Yasunaga, S. Kuhara, T. Shiba, M. Hattori, and H. Shinagawa. 2001. Complete genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res. 8:11–22.

56. Hayashi, T., H. Matsumoto, M. Ohnishi, and Y. Terawaki. 1993. Molecular analysis of a cytotoxin-converting phage, ϕCTX, of *Pseudomonas aeruginosa*: structure of the *attP*-cos-*ctx* region and integration into the serine tRNA gene. Mol. Microbiol. 7:657–667.

57. Hendrix, R. W., J. G. Lawrence, G. F. Hatfull, and S. Casjens. 2000. The origins and ongoing evolution of viruses. Trends Microbiol. 8:504–508.

58. Hendrix, R. W., M. C. M. Smith, R. N. Burns, and M. E. Ford. 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. Proc. Natl. Acad. Sci. USA 96:2192–2197.

59. Higgins, D. G., J. D. Thompson, and T. J. Gibson. 1996. Using CLUSTAL for multiple sequence alignments. Methods Enzymol. 266:383–402.

60. Hooper, I., and J. B. Egan. 1981. Coliphage 186 infection requires host initiation functions *dna*A and *dna*C. J. Virol. 40:599–601.

61. Hwang, E. S., and J. J. Scocca. 1990. Interaction of integration host factor from *Escherichia coli* with the integration region of the *Haemophilus influenzae* bacteriophage HP1. J. Bacteriol. 172:4852–4860.

62. Ilyina, T. V., and E. V. Koonin. 1992. Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. Nucleic Acids Res. 20: 3279–3285.

63. Jessop, L., T. Bankhead, D. Wong, and A. M. Segall. 2000. The amino terminus of bacteriophage lambda integrase is involved in protein–protein interactions during recombination. J. Bacteriol. 182:1024–1034.

64. Jin, S., Y.-C. Chen, G. E. Christie, and M. J. Benedik. 1996. Regulation of the *Serratia marcescens* extracellular nuclease: a positive control by a homolog of P2 Ogr encoded by a cryptic prophage. J. Mol. Biol. 256:264–278.

65. Juhala, R. J., M. E. Ford, R. L. Duda, A. Youlton, G. F. Hatfull, and R. W. Hendrix. 2000. Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. J. Mol. Biol. 299:27–51.

66. Julien, B., and R. Calendar. 1996. Bacteriophage PSP3 and ΦR73 activator proteins: analysis of promoter specificities. J. Bacteriol. 178:5668–5675.
67. Julien, B., and R. Calendar. 1995. Purification and characterization of the bacteriophage P4 δ protein. J. Bacteriol. 177:3743–3751.
68. Julien, B., P. Lefevre, and R. Calendar. 1997. The two Ogr-like domains of the Delta protein from bacteriophage P4 are required for activity. Virology 230:292–299.
69. Julien, B., D. Pountney, G. E. Christie, and R. Calendar. 1998. Mutational analysis of a satellite phage activator. Gene 223:129–134.
70. Kalionis, B., I. B. Dodd, and J. B. Egan. 1986. Control of gene expression in the P2-related temperate coliphages. III. DNA sequence of the major control region of phage 186. J. Mol. Biol. 191:199–209.
71. Kalionis, B., M. Pritchard, and J. B. Egan. 1986. Control of gene expression in the P2-related temperate coliphages. IV. Concerning the late control gene and control of its transcription. J. Mol. Biol. 191:211–220.
72. Kapfhammer, D., J. Blass, S. Evers, and J. Reidl. 2002. Vibrio cholerae phage K139: complete genome sequence and comparative genomics of related phages. J. Bacteriol. 184: 6592–6601.
73. Kim, S., and A. Landy. 1992. Lambda Int protein bridges between higher order complexes at two distant chromosomal loci *att*L and *att*R. Science 256:198–203.
74. King, R. A., D. L. Anders, and G. E. Christie. 1992. Site-directed mutagenesis of an amino acid residue in the bacteriophage P2 Ogr protein implicated in interaction with *Escherichia coli* RNA polymerase. Mol. Microbiol. 6:3313–3320.
75. Koonin, E. V., and T. V. Ilyina. 1993. Computer-assisted dissection of rolling circle DNA-replication. Biosystems 30:241–268.
76. Kwon, H. J., R. S. Tirumalai, A. Landy, and T. Ellenberger. 1997. Flexibility in DNA recombination: structure of the lambda integrase catalytic core. Science 276:126–131.
77. Lamont, I., H. Richardson, D. R. Carter, and J. B. Egan. 1993. Genes for the establishment and maintenance of lysogeny by the temperate coliphage 186. J. Bacteriol. 175:5286–5288.
78. Landy, A. 1989. Dynamic, strucural, and regulatory aspects of λ site-specific recombination. Annu. Rev. Biochem. 58:913–949.
79. Lee, T.-C., and G. E. Christie. 1990. Purification and properties of bacteriophage P2 *ogr* gene product. A procaryotic zinc-binding transcriptional activator. J. Biol. Chem. 265:7472–7477.
80. Lengyel, J., and R. Calendar. 1974. Control of bacteriophage P2 protein and DNA synthesis. Virology 57:305–313.
81. Lengyel, J. A., R. N. Goldstein, M. Marsch, and R. Calendar. 1974. Structure of the bacteriophage P2 tail. Virology 62:161–174.
82. Lennox, E. S. 1955. Transduction of linked genetic characters of the host by bacteriophage P1. Virology 1:190–206.
83. Lindahl, G. 1970. Bacteriophage P2: replication of the chromosome requires a protein which acts only on the genome that coded for it. Virology 42:522–533.
84. Linderoth, N. A., G. E. Christie, and E. Haggård-Ljungquist. 1993. Bacteriophage P2: physical, genetic and restriction map, pp. 1.62–1.69. *In* S. J. O'Brian (ed.) Genetic Maps, 6th edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
85. Linderoth, N. A., B. Julien, K. E. Flick, R. Calendar, and G. E. Christie. 1994. Molecular cloning and characterization of bacteriophage P2 genes *R* and *S* involved in tail completion. Virology 200:347–359.
86. Linderoth, N. A., R. Ziermann, E. Haggård-Ljungquist, G. E. Christie, and R. Calendar. 1991. Nucleotide sequence of the DNA packaging and capsid synthesis genes of bacteriophage P2. Nucleic Acids Res. 19:7207–7214.
87. Liu, T., and E. Haggård-Ljungquist. 1999. The transcriptional switch of bacteriophage WΦ, a P2-related but heteroimmune coliphage. J. Virol. 73:9816–9826.
88. Liu, T., S. K. Renberg, and E. Haggård-Ljungquist. 1997. Derepression of prophage P2 by satellite phage P4: cloning of the P4 ε gene and identification of its product. J. Virol. 71:4502–4508.
89. Liu, T., S. K. Renberg, and E. Haggård-Ljungquist. 1998. The E protein of satellite phage P4 acts as an anti-repressor by binding to the C protein of helper phage P2. Mol. Microbiol. 30:1041–1050.
90. Liu, Y., and E. Haggård-Ljungquist. 1996. Functional characterization of the P2 A initiator protein and its DNA cleavage site. Virology 216:158–164.
91. Liu, Y., and E. Haggård-Ljungquist. 1994. Studies of bacteriophage P2 DNA replication: localization of the cleavage site of the A protein. Nucleic Acids Res. 22:5204–5210.
92. Liu, Y., S. Saha, and E. Haggård-Ljungquist. 1993. Studies of bacteriophage P2 DNA replication. The DNA sequence of the *cis*-acting gene *A* and *ori* region and construction of a P2 mini-chromosome. J. Mol. Biol. 231:361–374.
93. Ljungquist, E., and L. E. Bertani. 1983. Properties and products of the cloned *int* gene of bacteriophage P2. Mol. Gen. Genet. 192:87–94.
94. Ljungquist, E., K. Kockum, and L. E. Bertani. 1984. DNA sequence of the repressor gene and operator region of bacteriophage P2. Proc. Natl. Acad. Sci. USA 81: 3988–3992.
95. Lurz, R., E. V. Orlova, D. Günther, P. Dube, A. Dröge, F. Weise, M. van Heel, and P. Tavares. 2001. Structural organization of the head-to-tail interface of a bacterial virus. J. Mol. Biol. 310:1027–1037.
96. Marvik, O. J., T. Dokland, R. H. Nøkling, E. Jacobsen, T. Larsen, and B. H. Lindqvist. 1995. The capsid size-determining protein Sid forms an external scaffold on phage P4 procapsids. J. Mol. Biol. 251:59–75.
97. Marvik, O. J., E. Jacobsen, T. Dokland, and B. H. Lindqvist. 1994. Bacteriophage P2 and P4 morphogenesis: assembly precedes proteolytic processing of the capsid proteins. Virology 205:51–65.
98. Marvik, O. J., P. Sharma, T. Dokland, and B. H. Lindqvist. 1994. Bacteriophage P2 and P4 assembly: alternative scaffolding proteins regulate capsid size. Virology 200: 702–714.
99. McClelland, M., K. E. Sanderson, J. Spieth, S. W. Clifton, P. Latreille, L. Courtney, S. Porwollik, J. Ali, M. Dante, F. Du, S. Hou, D. Layman, S. Leonard, C. Nguyen, K. Scott, A. Holmes, N. Grewal, E. Mulvaney, E. Ryan, H. Sun,

L. Florea, W. Miller, T. Stoneking, M. Nhan, R. Waterston, and R. K. Wilson. 2001. Complete genome sequence of *Salmonella enterica* serovar *Typhimurium* LT2. Nature 413:852–856.

100. Mirold, S., W. Rabsch, H. Tschäpe, and W.-D. Hardt. 2001. Transfer of the *Salmonella* type III effector *sopE* between unrelated phage families. J. Mol. Biol. 312:7–16.
101. Mosig, G., S. Yu, H. Myung, E. Haggård-Ljungquist, L. Davenport, K. Carlsson, and R. Calendar. 1997. A novel mechanism of virus–virus interactions: bacteriophage P2 Tin protein inhibits phage T4 DNA synthesis by poisoning the T4 single-stranded DNA binding protein, gp32. Virology 230:72–81.
102. Myung, H., and R. Calendar. 1995. The *old* nuclease of bacteriophage P2. J. Bacteriol. 177:497–501.
103. Nakayama, K., S. Kanaya, M. Ohnishi, Y. Terawaki, and T. Hayashi. 1999. The complete nucleotide sequence of ΦCTX, a cytotoxin-converting phage of *Pseudomonas aeruginosa*: implications for phage evolution and horizontal transfer via bacteriophages. Mol. Microbiol. 31: 399–419.
104. Nakayama, K., K. Takashima, H. Ishihara, T. Shinomiya, M. Kageyama, S. Kanaya, M. Ohnishi, T. Murata, H. Mori, and T. Hayashi. 2000. The R-type pyocin of *Pseudomonas aeruginosa* is related to P2 phage, and the F-type is related to lambda phage. Mol. Microbiol. 38:213–231.
105. Nesper, J., J. Blass, M. Fountoulakis, and J. Reidl. 1999. Characterization of the major control region of *Vibrio cholerae* bacteriophage K139: immunity, exclusion, and integration. J. Bacteriol. 181:2902–2913.
106. Neufing, P. J., K. E. Shearwin, J. Camerotto, and J. B. Egan. 1996. The CII protein of bacteriophage 186 establishes lysogeny by activating a promoter upstream of the lysogenic promoter. Mol. Microbiol. 21:751–761.
107. Neufing, P. J., K. E. Shearwin, and J. B. Egan. 2001. Establishing lysogenic transcription in the temperate coliphage 186. J. Bacteriol. 183:2376–2379.
108. Nilsson, A. S., and E. Haggård-Ljungquist. 2001. Detection of homologous recombination among bacteriophage P2 relatives. Mol. Phylogenet. Evol. 21:259–269.
109. Nilsson, A. S., J. L. Karlsson, and E. Haggård-Ljungquist. 2004. Site-specific recombination links the evolution of P2-like coliphages and pathogenic enterobacteria. Mol. Biol. Evol. 21:1–13.
110. Nunes-Düby, S. E., H. J. Kwon, R. S. Tirumalai, T. Ellenberger, and A. Landy. 1998. Similarities and differences among 105 members of the Int family of site-specific recombinases. Nucleic Acids Res. 26:391–406.
111. Ochman, H., and R. K. Selander. 1984. Standard reference strains of *Escherichia coli* from natural populations. J. Bacteriol. 157:690–693.
112. Odegrip, R., and E. Haggård-Ljungquist. 2001. The two active-site tyrosine residues of the A protein play non-equivalent roles during initiation of rolling circle replication of bacteriophage P2. J. Mol. Biol. 308:147–163.
113. Odegrip, R., S. Schoen, E. Haggård-Ljungquist, K. Park, and D. K. Chattoraj. 2000. The interaction of bacteriophage P2 B protein with *Escherichia coli* DnaB helicase. J. Virol. 74:4057–4063.

113a. Odegrip, R. 2001. Initiation of rolling-circle replication. PhD thesis, Stockholm University, Stockholm, Sweden.

114. Perna, N. T., G. Plunkett III, V. Burland, B. Mau, J. D. Glasner, D. J. Rose, G. F. Mayhew, P. S. Evans, J. Gregor, H. A. Kirkpatrick, G. Pósfal, J. Hackett, S. Klink, A. Boutin, Y. Shao, L. Miller, E. J. Grotbeck, N. W. Davis, A. Lim, E. T. Dimalanta, K. D. Potamousis, J. Apodaca, T. S. Anantharaman, J. Lin, G. Yen, D. C. Schwartz, R. A. Welch, and F. R. Blattner. 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature 409:529–533.
115. Pontarollo, R. A., C. R. Rioux, and A. A. Potter. 1997. Cloning and characterization of bacteriophage-like DNA from *Haemophilus somnus* homologous to phages P2 and HP1. J. Bacteriol. 179:1872–1879.
116. Portelli, R., I. B. Dodd, Q. Xue, and J. B. Egan. 1998. The late-expressed region of the temperate coliphage 186 genome. Virology 248:117–130.
117. Pountney, D. L., R. P. Tiwari, and J. B. Egan. 1997. Metal- and DNA-binding properties and mutual anaysis of the transcription activating factor, B, of coliphage 186: a prokaryotic C4 zinc-finger protein. Protein Sci. 6:892–902.
118. Ptashne, M. 1992. A genetic switch: phage λ and Higher Organisms, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
119. Reed, M. R., F. E. Shearwin, L. M. Pell, and J. B. Egan. 1997. The dual role of Apl in prophage induction of coliphage 186. Mol. Microbiol. 23:669–681.
120. Renberg-Eriksson, S. K., A. Ahlgren-Berg, J. DeGrooth, and E. Haggård-Ljungquist. 2001. Characterization of the developmental switch region of bacteriophage P2 Hy *dis*. Virology 290:199–210.
121. Renberg-Eriksson, S. K., T. Liu, and E. Haggård-Ljungquist. 2000. Interacting interfaces of the P4 antirepressor E and the P2 immunity repressor C. Mol. Microbiol. 36:1148–1155.
122. Rice, P. A., S.-W. Yang, K. Mizuuchi, and H. A. Nash. 1996. Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. Cell 87:1295–1306.
123. Richardson, H., and J. B. Egan. 1989. Control of gene expression in the P2-related temperate coliphage 186. VI. Sequence analysis of the early lytic region. J. Mol. Biol. 206:251–255.
124. Richardson, H., and J. B. Egan. 1989. DNA replication studies with coliphage 186. II. Derepression of host replication by a 186 gene. J. Mol. Biol. 206:59–68.
125. Rishovd, S., A. Holzenburg, B. V. Johansen, and B. H. Lindqvist. 1998. Bacteriophage P2 and P4 morphogenesis: structure and function of the connector. Virology 245:11–17.
126. Rishovd, S., and B. H. Lindqvist. 1992. Bacteriophage P2 and P4 morphogenesis: protein processing and capsid size determination. Virology 187:548–554.
127. Rishovd, S., O. J. Marvik, E. Jacobsen, and B. H. Lindqvist. 1994. Bacteriophage P2 and P4 morphogenesis: identification and characterization of the portal protein. Virology 200:744–751.
128. Saha, S., E. Haggård-Ljungquist, and K. Nordström. 1989. Activation of prophage P4 by the P2 Cox protein and

the sites of action of the Cox protein on the two phage genomes. Proc. Natl. Acad. Sci. USA 86:3973–3977.

129. Saha, S., E. Haggård-Ljungquist, and K. Nordström. 1987. The Cox protein of bacteriophage P2 inhibits the formation of the repressor protein and autoregulates the early operon. EMBO J. 6:3191–3199.

130. Saha, S., B. Lundqvist, and E. Haggård-Ljungquist. 1987. Autoregulation of bacteriophage P2 repressor. EMBO J. 6:809–814.

131. Sauer, B., R. Calendar, E. Ljungquist, E. Six, and M. G. Sunshine. 1982. Interaction of satellite phage P4 with phage 186 helper. Virology 116:523–534.

132. Shearwin, K. E., A. M. Brumby, and J. B. Egan. 1998. The Tum protein of coliphage 186 is an antirepressor. J. Biol. Chem. 273:5708–5715.

133. Shearwin, K. E., I. B. Dodd, and J. B. Egan. 2002. The helix–turn–helix motif of the coliphage 186 immunity repressor binds to two distinct recognition sequences. J. Biol. Chem. 277:3186–3194.

134. Shearwin, K. E., and J. B. Egan. 1996. Purification and self-association equilibria of the lysis–lysogeny switch proteins of coliphage 186. J. Biol. Chem. 271:11525–11531.

135. Sinsheimer, R. L., C. A. Hutchison, and B. H. Lindqvist. 1967. Bacteriophage ΦX174: viral functions, pp. 175–192. *In* J. S. Colter, and W. Parachych (eds.) The Molecular Biology of Viruses. Academic Press, New York.

136. Sivaprasad, A. V., R. Jarvinen, A. Puspurs, and J. B. Egan. 1990. DNA replication studies with coliphage 186. III. A single phage gene is required for phage 186 replication. J. Mol. Biol. 213:449–463.

137. Six, E. W., M. G. Sunshine, J. Williams, E. Haggård-Ljungquist, and B. H. Lindqvist. 1991. Morphopoietic switch mutations in bacteriophage P2. Virology 182:34–46.

138. Slettan, A., K. Gebhardt, E. Kristiansen, N.-K. Birkeland, and B. H. Lindqvist. 1992. *Escherichia coli* K-12 and B contain functional bacteriophage P2 *ogr* genes. J. Bacteriol. 174:4094–4100.

139. Souza, L., R. Calendar, E.W. Six, and B. H. Lindqvist. 1977. A *trans*activation mutant of satellite phage P4. Virology 81:81–90.

140. Sunshine, M. G., and B. Sauer. 1975. A bacterial mutation blocking P2 phage late gene expression. Proc. Natl. Acad. Sci. USA 72:2770–2774.

141. Swofford, D. L. 1999. PAUP*: Phylogenetic Analysis Using Parsimony (* and Other Methods), 4.0 edn. Sinauer Associates, Sunderland, Mass.

142. Temple, L. M., S. L. Forsburg, R. Calendar, and G. E. Christie. 1991. Nucleotide sequence of the genes encoding the major tail sheath and tail tube proteins of bacteriophage P2. Virology 181:353–358.

143. Tessman, E. S. 1966. Mutants of bacteriophage S13 blocked in infectious DNA synthesis. J. Mol. Biol. 17:218–236.

144. Tessman, E. S., M. T. Borras, and I. L. Sun. 1971. Superinfection in bacteriophage S13 and determination of the number of bacteriophage particles which can function in an infected cell. J. Virol. 8:111–120.

145. Valpuesta, J. M., and J. L. Carrascosa. 1994. Structure of viral connectors and their function in bacteriophage assembly and DNA packaging. Q. Rev. Biophys. 27:107–155.

146. Valpuesta, J. M., J. J. Fernández, J. M. Carazo, and J. L. Carrascosa. 1999. The three-dimensional structure of a DNA translocating machine at 10 Å resolution. Structure Fold Des. 7:289–296.

147. Van Bokkelen, G. B., E. C. Dale, C. Halling, and R. Calendar. 1991. Mutational analysis of a bacteriophage P4 late promoter. J. Bacteriol. 173:37–45.

148. Van der Mei, D., J. Zandberg, and H. S. Jansz. 1972. The effect of chloramphenicol on synthesis of ΦX174-specific proteins and deletion of the cistron A protein. Biochim. Biophys. Acta 287:312–321.

149. Van Mansfeld, A. D. M., H. A. A. M. Van Teeffelen, P. D. Baas, and H. S. Jansz. 1986. Two juxtaposed tyrosyl-OH groups participate in ΦX174 gene A protein catalysed cleavage and ligation of DNA. Nucleic Acids Res. 14:4229–4238.

150. Wang, I.-N., D. L. Smith, and R. Young. 2000. Holins: the protein clocks of bacteriophage infections. Annu. Rev. Microbiol. 54:799–825.

151. Williams, B. J., M. Golomb, T. Phillips, J. Brownlee, M. V. Olson, and A. L. Smith. 2002. Bacteriophage HP2 of *Haemophilus influenzae*. J. Bacteriol. 184:6893–6905.

152. Williams, S. G., and J. B. Egan. 1994. DNA replication studies with coliphage 186: the involvement of the *Eschrichia coli* DnaA protein in 186 replication is indirect. J. Bacteriol. 176:6039–6044.

153. Xue, Q., and J. B. Egan. 1995. Tail sheath and tail tube genes of the temperate coliphage 186. Virology 212:218–221.

154. Young, R. 2002. Bacteriophage holins: deadly diversity. J. Mol. Microbiol. Biotechnol. 4:21–36.

155. Young, R. 1992. Bacteriophage lysis: mechanism and regulation. Microbiol. Rev. 56:430–481.

156. Yu, A., V. Barreiro, and E. Haggård-Ljungquist. 1994. Regulation of *int* gene expression in bacteriophage P2. J. Virol. 68:4220–4226.

157. Yu, A., L. E. Bertani, and E. Haggård-Ljungquist. 1989. Control of prophage integration and excision in bacteriophage P2: nucleotide sequences of the *int* gene and *att* sites. Gene 80:1–12.

158. Yu, A., and E. Haggård-Ljungquist. 1993. Characterization of the binding sites of two proteins involved in the bacteriophage P2 site-specific recombination system. J. Bacteriol. 175:1239–1249.

159. Yu, A., and E. Haggård-Ljungquist. 1993. The Cox protein is a modulator of directionality in bacteriophage P2 site-specific recombination. J. Bacteriol. 175:7848–7855.

160. Ziermann, R., B. Bartlett, R. Calendar, and G. E. Christie. 1994. Functions involved in bacteriophage P2-induced host cell lysis and identification of a new tail gene. J. Bacteriol. 176:4974–4984.

161. Ziermann, R., and R. Calendar. 1990. Characterization of the *cos* site of bacteriophages P2 and P4. Gene 96:9–15.

26

The Satellite Phage P4

GIANNI DEHÒ
DANIELA GHISOTTI

P4, the Satellite Phage, the Plasmid

In the early 1960s Erich Six isolated a temperate bacteriophage, P4, that formed plaques only on *Escherichia coli* strains lysogenic for P2 or a P2-related phage (90, 91). Before long, it was established that P4 depended on the P2 helper phage for all morphopoietic and lysis "late" functions, and that the head and tail proteins of P4 virions were encoded by P2 (51, 92). Interestingly, the smaller P4 genome, encapsidated into a smaller head (figure 26-1), did not bear any substantial similarity to P2 DNA and was independent of the helper for DNA replication and host lysogenization (68, 91, 92). The functional and structural unrelatedness of the two replicons clearly indicated that P4 was not a defective P2 but an independent entity and the prototype of "satellite viruses," genetic elements that exploit unrelated viruses for viral propagation.

Later it emerged that, in the absence of the helper genome, P4 could be maintained not only as an integrated prophage but also as a multicopy plasmid (28, 44, 68, 75). Because of its diverse modes of propagation (see figure 26-2), P4 may be considered as an integrative plasmid that has evolved the potential for horizontal transfer by a very specialized phage-mediated transduction mechanism. Like P4, other bacteriophages such as the filamentous phages and the temperate phages P1 and N15, may be stably maintained as autonomously replicating plasmids in the bacterial host. These natural phasmids (phage-plasmid) make the boundary between plasmids and phages less definite and contribute to our understanding of viruses as part of a large pool of mobile genetic information that can be exchanged among living cells.

P4 is no longer a unique example of satellite phages: in addition to retronphage ΦR73, a P4-like phage isolated from *E. coli* (52, 98), a satellite plasmid pSSVx/helper-phage SSV2 pair, functionally resembling the P4/P2 system, has been discovered in the archaeon *Sulfolobus* (4).

The Genome

A P4 virion is made of a linear, double-stranded, cohesive-ended DNA molecule 11.6 kb long that is encapsidated into a tailed, icosahedral protein head (figure 26-1). The genome organization of P4 is diagrammed in figure 26-3A and the known genes and sites are detailed in table 26-1. Upon infection of the *E. coli* host, P4 DNA circularizes through its cohesive ends. The genetic vegetative map of P4 is circular (8, 27) and reflects the physical structure of the replicating DNA molecule, whereas the integrated prophage genome is a circular permutation of the mature DNA (17a).

All the functions required for lysogenic, lytic, and plasmid development are located in the right 80% of the genome and include the origin of replication, *ori1* (63), the two main α and *sid* operons which are transcribed divergently from the *ori1* site (26, 30, 50), the prophage integration *att* site (17a, 78), and the *int* (integrase) gene in a monocistronic operon located to the left of *att*.

As in many episomal elements, a "nonessential region" that carries functions not related to the P4 life cycle (*gop*, β, *cII*; table 26-1) is found adjacent to the *att–int* integration module (17a, 41). These genes, organized in the two constitutively expressed operons to the left of *att–int*, may be deleted without affecting lysogenic, lytic, or plasmid development (41, 58).

Phages P4 and ΦR73 exhibit more than 95% sequence identity in the essential region between *crr* and the right *cos* site, with the exception of gene δ (31%) and the immunity region (85%; see below). On the contrary, in phage ΦR73 the nonessential region to the left of *att–int* is largely deleted. To the right of the integration module ΦR73 carries a retron, a genetic element encoding reverse transcriptase that is responsible for the synthesis of a peculiar single-stranded DNA–RNA chimeric molecule called msDNA (52, 98). The divergent genetic organization of the nonessential regions suggests that phages P4 and ΦR73

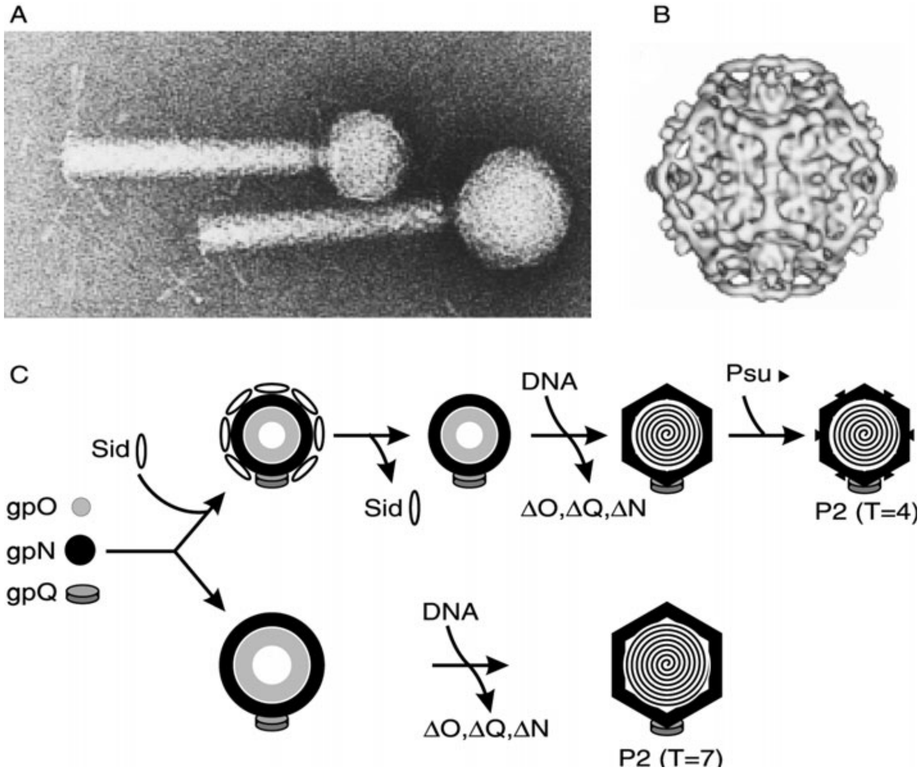

A
B
C
Sid
gpO
gpN
gpQ
Sid
DNA
ΔO,ΔQ,ΔN
Psu ▸
P2 (T=4)
DNA
ΔO,ΔQ,ΔN
P2 (T=7)

Figure 26-1 Structure and morphogenesis of satellite bacteriophage P4 and its helper P2 virions. A: P4 (top) and P2 (bottom) virions. Negative stain with phosphotungstate. The tail structures are about 135 nm long. The P4 and P2 heads are, respectively, 45 nm and 62 nm in diameter. Electron micrograph by Robley C. Williams, University of California, Berkeley. From (43), with permission. B: Structure of the P4 procapsid with the Sid external scaffold (dark ribbon). Reconstruction of P4 procapsids at 2.6 nm resolution. The procapsids, approximately 40 nm in diameter, were obtained by coexpressing in vivo P2 gpN and P4 Sid. From (35), with permission. C: Comparison of the assembly pathways of P4 and P2 capsids. Δ indicates a cleaved protein. Redrawn from (34), with permission.

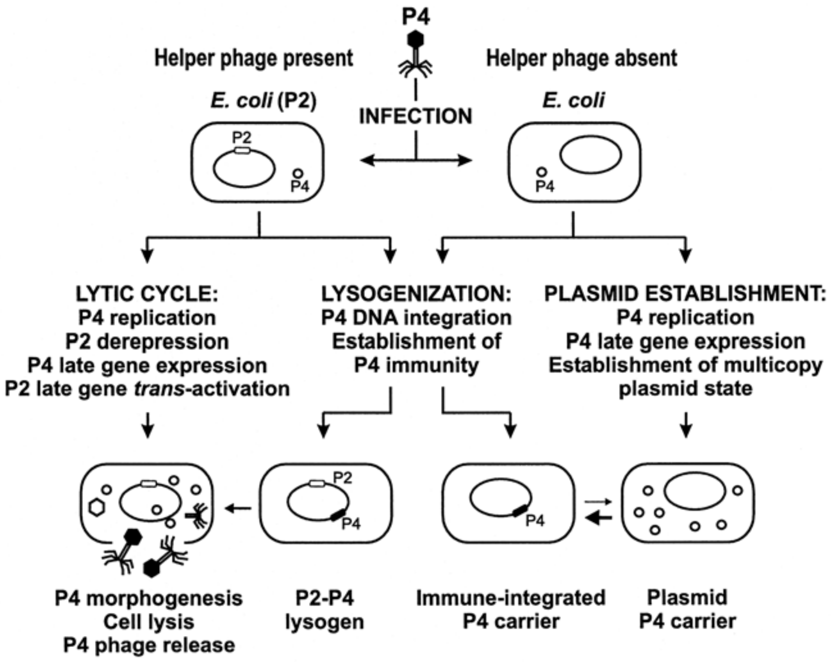

P4
Helper phage present
Helper phage absent
E. coli (P2)
INFECTION
E. coli
P2
P4
P4
LYTIC CYCLE:
P4 replication
P2 derepression
P4 late gene expression
P2 late gene trans-activation
LYSOGENIZATION:
P4 DNA integration
Establishment of
P4 immunity
PLASMID ESTABLISHMENT:
P4 replication
P4 late gene expression
Establishment of multicopy
plasmid state
P2
P4
P4
P4 morphogenesis
Cell lysis
P4 phage release
P2-P4
lysogen
Immune-integrated
P4 carrier
Plasmid
P4 carrier

Figure 26-2 Phage P4 life cycle. See the text for explanation. Redrawn from (69), with permission.

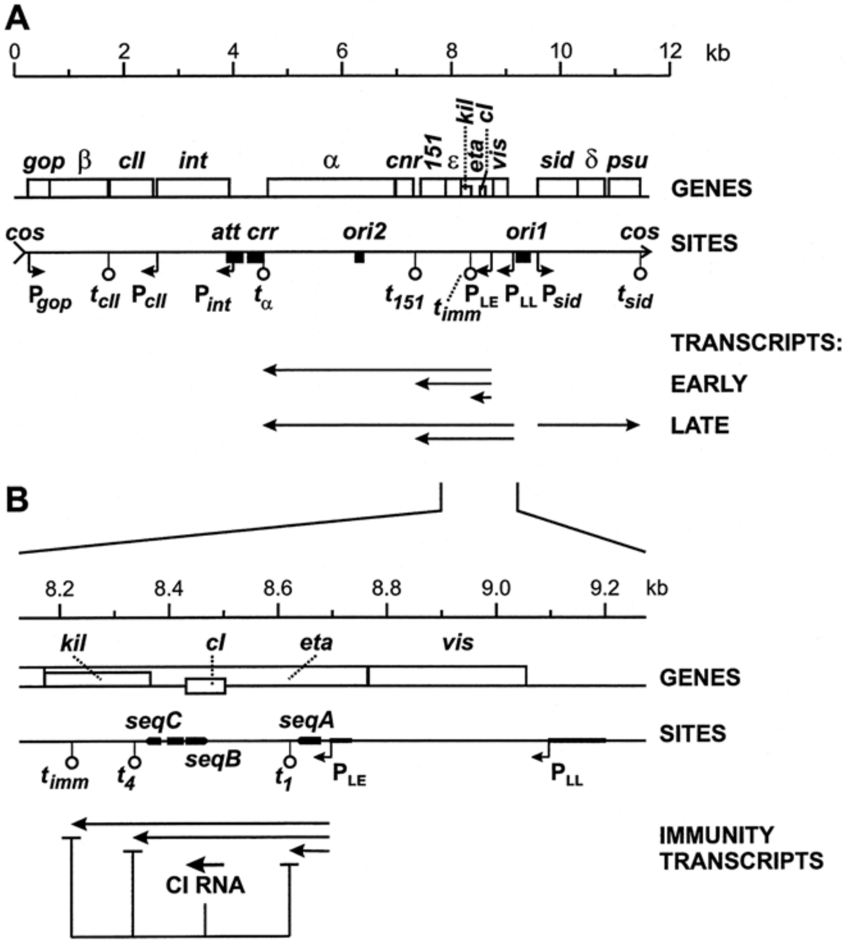

A
0
2
4
6
8
10
12 kb
gop
β
cII
int
α
cnr
151
ε
kil
eta
cI
vis
sid
δ
psu
GENES
cos
att crr
ori2
ori1
cos
SITES
Pgop
tcII
PcII
Pint
tα
t151
timm
PLE
PLL
Psid
tsid
TRANSCRIPTS:
EARLY
LATE
B
8.2
8.4
8.6
8.8
9.0
9.2 kb
kil
cI
eta
vis
GENES
seqC
seqA
seqB
SITES
timm
t4
t1
PLE
PLL
IMMUNITY
TRANSCRIPTS
CI RNA

Figure 26-3 Phage P4 genome. A: Physical and genetic map of P4. B: P4 immunity region. Coordinates are from the annotated complete nucleotide sequence of P4 (47) (GenBank accession number X51522). Transcription start and termination sites are indicated by bent arrows and hanging open circles, respectively. The arrows beneath the maps in panel A and B indicate the early, late and immune phase transcripts from the two main operons. For other explanations see table 26-1 and the text. Redrawn from (85), with permission.

evolved via independent illegitimate integration–excision cycles.

The α operon, which is transcribed leftwards from two promoters, P_{LE} and P_{LL} (30), encodes both genes required for P4 lytic and plasmid propagation (α and *cnr*, replication; ε, helper prophage derepression), and the immunity determinants, required to prevent P4 prophage replication. Therefore, control of α operon expression is crucial for the establishment and maintenance of the different developmental phases and involves the sophisticated mechanisms described below.

The *sid* operon codes for regulatory and morphogenetic proteins involved in plasmid and/or lytic development (26): the positive regulator, δ, which can both activate the satellite and *trans*-activate the helper late operons (26, 30, 95), and two proteins with a role in P4 head morphogenesis. The latter are Sid, which forms a head external scaffold and determines the small size of the P4 capsid (1, 74, 89), and Psu, a bifunctional protein that suppresses transcription termination at Rho-dependent terminators (polarity suppressor; 64, 67, 88) and helps to stabilize the viral particle (capsid "decoration" protein; 54, 55). The *sid* operon is transcribed late after infection from the P_{sid} promoter, which is activated by P4- and P2-encoded positive regulators (gpδ and Ogr, respectively; 26, 46).

The Replicon

Autonomous P4 DNA replication occurs both in the lytic cycle and during plasmid propagation, and is completely independent of helper phage functions. Immediately after infection, a burst of P4 replication occurs and a large number of P4 DNA molecules (>100) accumulate in the host cell. On the contrary, in the plasmid state DNA replication is controlled so as to maintain a constant copy

Table 26-1 Bacteriophage P4 Genes and Functions

Gene or Site	Gene Product and/or Function Encoded
cos	Cohesive ends 19 nucleotides long
P_{gop}	Promoter of the *gop*-β *operon*
gop	Causes host cell killing in the absence of β
β	Inhibits *gop* killing
t_{cII}	Rho-independent transcription termination site
cII	Function unknown. Mutants kill the host cell
P_{cII}	Promoter of the *cII* gene
int	Integrase
P_{int}	Promoter of the *int* gene
att	Site for integrative recombination
crr	Required in *cis* for replication of both *oriI* and *oriII* replicons
t_{α}	Rho-independent transcription termination site
α	Essential for replication; primase, helicase, *ori1*, and *crr* recognition and binding
ori2	With *crr* supports α-dependent *ori1*-independent replication (*oriII* replicon)
cnr	Controls DNA replication and plasmid copy number; interacts with α protein
t_{151}	Putative transcription terminator
orf151	Function unknown
ε	Derepression of the P2 helper prophage
kil	Kills the bacterial host if overexpressed
t_{imm}	Rho-dependent transcription terminator. Elicits strong transcription termination from P_{LE} when the CI RNA is present
t_4	Rho-independent, CI RNA-dependent transcription termination site
cI	Prophage immunity. Encodes the CI RNA
t_1	Rho-independent transcription termination site
P_{LE}	Constitutive promoter
eta	Function of gene product unknown. Its translation prevents transcription termination from P_{LL}
vis	Binds P_{LL}, P_{sid}, and *att*; negative regulator of P_{LL}; stimulates P_{sid}; excisionase
P_{LL}	Late promoter; positively regulated by P4 gpδ and P2 Ogr and Cox; negatively regulated by Vis
ori1	Origin of DNA replication
P_{sid}	Late promoter; positively regulated by P4 δ and P2 Ogr, stimulated by Vis
sid	Small head determination; procapsid external scaffold
δ	P4 and P2 late promoter activator
psu	Polarity suppression; capsid decoration protein
t_{sid}	Rho-independent transcription termination site

number of about 40 P4 genomes per bacterial chromosome (2). This control is exerted both by limiting the expression of the P4 replication protein, gpα, and by directly controlling the activity of gpα in DNA replication initiation (see below). P4 DNA replication starts at a unique point, *ori1*, and proceeds bidirectionally in the θ form (63).

The P4 DNA replication system is peculiar among other phage or plasmid replication systems in that P4 is largely independent of bacterial functions for replication initiation; in fact, P4 encodes the multifunctional α protein that exhibits primase and helicase activity, and specifically binds DNA sequences at the replication origin (36, 108). Accordingly, host functions involved in DNA replication initiation, such as DnaA (initiator), DnaB (helicase), DnaC (DnaB complex), and DnaG (primase), as well as the Rep helicase, which is involved in the chromosome replication forks progression, are not required in vivo (6, 10, 33, 62, 68). In vitro, replication of P4 DNA molecules has been obtained using combined mixtures of gpα, DNA polymerase III holoenzyme, single-stranded DNA-binding protein, DNA gyrase, and topoisomerase (33).

The P4 α Protein

The P4 α gene, essential for P4 DNA replication, encodes a 777 amino acid polypeptide. Deletion and point mutants have been used to dissect the functional domains of gpα by testing their ability to sustain phage replication in vivo and/or assaying in vitro the activity of the mutant proteins. The modular structure of gpα emerged from this analysis, with primase and helicase activities arranged in distinct, separable domains (figure 26-4A).

The primase activity is located in the N-terminal half of the α protein. A potential metal-binding region, with two CXXC clusters and the EGYATA motif, common to primases of conjugative plasmids, is essential for P4 replication in vivo and primase activity in vitro. A potential Mg^{2+}-binding region was also found (96, 108, 109).

The helicase activity, characterized by the type A nucleotide binding motif, requires the middle and C-terminal parts of gpα. Point mutations in the type A motif are defective in phage propagation in vivo and in helicase activity in vitro (96, 109, 111).

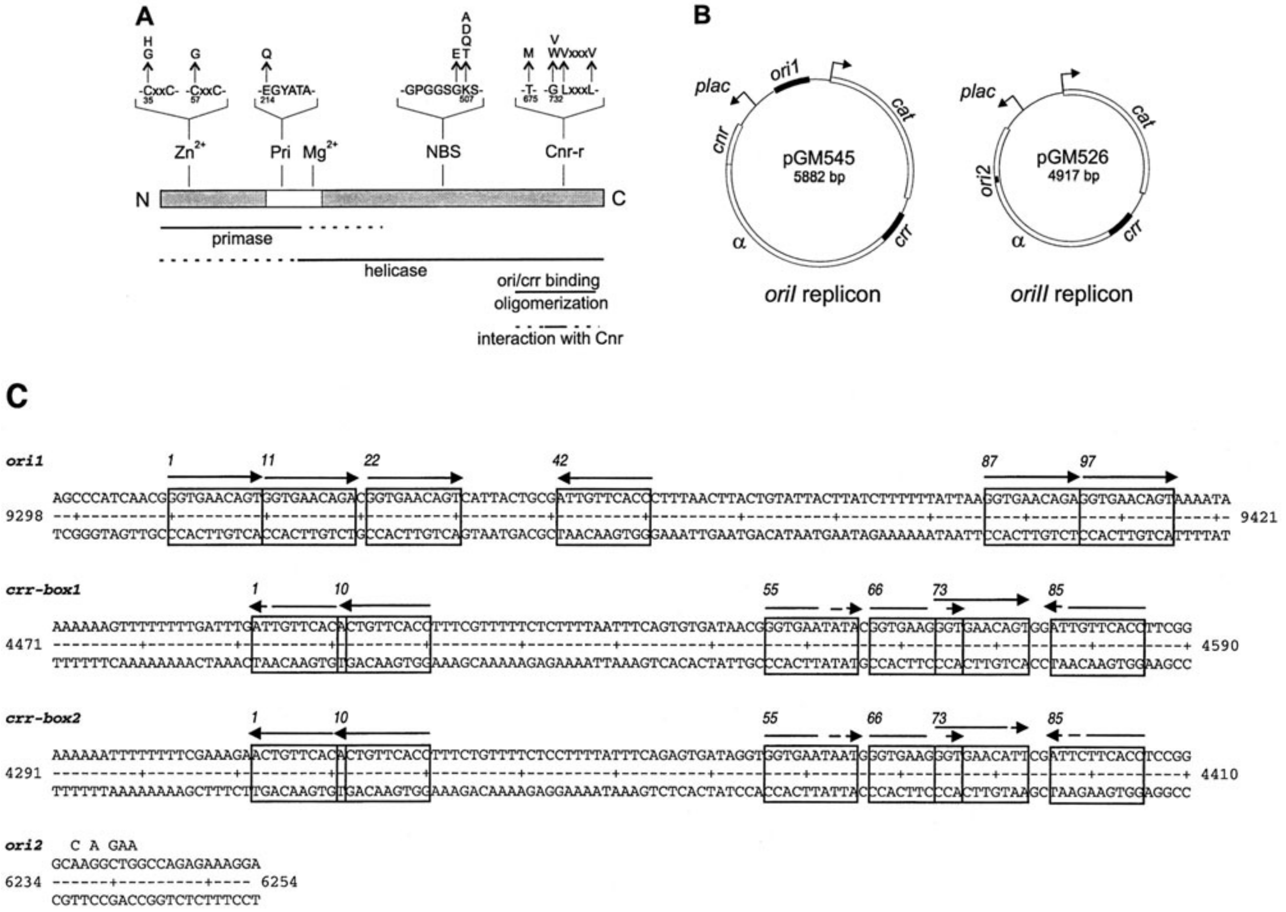

A
H
G
G
Q
A
D
Q
E T
M
V
WVxxxV
-CxxC- 35
-CxxC- 57
-EGYATA- 214
-GPGGSGKS- 507
-T- 675
-G LxxxL- 732
Zn2+
Pri
Mg2+
NBS
Cnr-r
N
C
primase
helicase
ori/crr binding
oligomerization
interaction with Cnr
B
plac
ori1
cat
cnr
pGM545
5882 bp
α
crr
oriI replicon
plac
cat
ori2
pGM526
4917 bp
α
crr
oriII replicon
C
ori1
1
11
22
42
87
97
AGCCCATCAACGGGTGAACAGTGGTGAACAGACGGTGAACAGTCATTACTGCGATTGTTCACCCTTTAACTTACTGTATTACTTATCTTTTTTATTAAGGTGAACAGAGGTGAACAGTAAAATA
9298
9421
TCGGGTAGTTGCCCACTTGTCACCACTTGTCTGCCACTTGTCAGTAATGACGCTAACAAGTGGGAAATTGAATGACATAATGAATAGAAAAAATAATTCCACTTGTCTCCACTTGTCATTTTAT
crr-box1
1
10
55
66
73
85
AAAAAAGTTTTTTTTGATTTGATTGTTCACACTGTTCACCTTTCGTTTTTCTCTTTTAATTTCAGTGTGATAACGGGTGAATATACGGTGAAGGGTGAACAGTGGATTGTTCACCTTCGG
4471
4590
TTTTTTCAAAAAAAACTAAACTAACAAGTGTGACAAGTGGAAAGCAAAAAGAGAAAATTAAAGTCACACTATTGCCCACTTATATGCCACTTCCCACTTGTCACCTAACAAGTGGAAGCC
crr-box2
1
10
55
66
73
85
AAAAAATTTTTTTTCGAAAGAACTGTTCACACTGTTCACCTTTCTGTTTTCTCCTTTTATTTCAGAGTGATAGGTGGTGAATAATGGGTGAAGGGTGAACATTCGATTCTTCACCTCCGG
4291
4410
TTTTTTAAAAAAAAGCTTTCTTGACAAGTGTGACAAGTGGAAAGACAAAAGAGGAAAATAAAGTCTCACTATCCACCACTTATTACCCACTTCCCACTTGTAAGCTAAGAAGTGGAGGCC
ori2
C A GAA
GCAAGGCTGGCCAGAGAAAGGA
6234
6254
CGTTCCGACCGGTCTCTTTCCT

Figure 26-4 The essential elements for phage P4 replication. A: Fuctional domains of the P4 α protein. The gray bar represents the α protein; its functional domains are indicated by the continous and dotted lines below. Highlighted in white is a 120 amino acid region exhibiting sequence similarity to primases of IncI and IncP plasmids. The amino acid consensus sequences of conserved motifs are reported on top, and the amino acid changes affecting the different activities are listed over the arrows. Pri, primase motif conserved among several prokaryotic primases; Zn^{2+}, metal binding motif found in other prokaryotic primases and DNA-repair proteins; Mg^{2+}, potential metal binding site, has similarity with other DNA and RNA polymerases; NBS, type A nucleotide binding site found in other helicases of small DNA and RNA viruses; the Cnr-r site, defined by Cnr resistance mutations, interacts with the Cnr-protein. From (109), with permission. B: Replicon oriI and *oriII*. Plasmids pGM545 and pGM526 carry the essential P4 regions of the two P4 replicons, ligated to the chloramphenicol resistance gene. Replicon *oriI*: *ori1* and *crr* sites, and *cnr* and α genes; replicon *oriII*: *ori2* and *crr* sites, and α gene. The genes are expressed under *plac* control. From (103), with permission. **C:** The minimal *ori1* sequence and the two direct repeats of *crr* are reported. The type I iterons are boxed. The arrows indicate the orientation and the number above the arrows indicate the position of the first base of each iteron, relative to the first iteron from the left. For *ori2*, the sequence of the minimal 22 bp region is reported. Above the sequence is the five-base substitution that inactivates *ori2* (D. Ghisotti, unpublished data).

Other functional domains are located in the C-terminal end of gpα that are retained by the truncated C-terminal peptide: (i) a DNA-binding domain, which specifically binds to *ori1* and *crr*, and (ii) a homodimerization domain (104, 109). The crystal structure of the α protein C-terminal origin-binding domain, recently solved at 0.295 nm resolution, reveals an overall fold of the winged helix subfamily of helix–turn–helix DNA-binding proteins, forming homodimers with pseudo 2-fold symmetry (107). In addition, a cluster of mutations that make P4 replication insensitive to the negative regulation exerted by the Cnr protein (see below), maps to the 3′ end of the α gene, suggesting that the gpα–Cnr interaction domain is also located in the C-terminal part of the protein (110). Interestingly, all mutations that make gpα insensitive to Cnr protein map to the dimerization interface (107).

The oriI Replicon

The P4 origin of replication, *oriI*, is bipartite, composed of the *ori1* and *crr* sites (figure 26-4B, C). Both *ori1* and *crr* contain several direct and inverted repeats of a decameric sequence, the type I iterons (GGTGAACAGA/T), which are bound by the α protein (36, 63, 103, 108). Moreover, both sites contain AT-rich stretches.

In the P4 genome, the *ori1* and *crr* sites lie about 4500 bp apart. The spacing between *crr* and *ori1* can be reduced to less than 100 bp without affecting replication (36). However, the relative orientation of the two sites is essential (20). Moreover, *ori1* cannot substitute for *crr* (103), indicating that they have different roles in replication.

The minimal *ori1* site was limited to a 123 bp region, which contains six type I iterons regularly spaced with helical periodicity and a central 35 bp AT-rich region (103) (figure 26-4C). Both the number and the spacing of the type I iterons are essential for replication (103). Thus initiation of P4 replication might occur in a similar way to the other iteron-containing origins, such as *E. coli oriC* or the phage P1 *oriR* (11, 12, 106), in which binding of the initiator protein causes DNA bending and wrapping around a core of the initiation proteins. In P4, several α proteins bound to the *ori1* site in a regular arrangement might constitute a nucleoprotein complex competent for replication. The central AT-rich region might be involved in specific unwinding and primer synthesis.

The *crr* site is formed by two 120 bp AT-rich repeats, each containing five type I iterons (figure 26-4C). However, a single 120 bp *crr* repeat is sufficient to promote replication, although the efficiency is reduced about 10-fold (36, 103). Type I iterons in *crr* do not show helical periodicity. The α protein binds to these iterons, but initiation of replication from *crr* has never been observed, either in vivo or in vitro (33, 62). *crr* is therefore not a replication start point but is essential for replication initiation at *ori1*.

In the presence of the α protein, looping of P4 DNA molecules containing *ori1* and *crr* was observed (33, 108, 109). This suggests that α proteins, bound to *ori1* and *crr*, might interact with each other. Interaction between α-*ori1* and α-*crr* might be required for the formation of an active replication initiation complex, possibly by the rearrangement of the α-*ori1* complex to make it competent for initiation of bidirectional replication.

Control of P4 Replication

Many plasmids encode feedback systems that negatively control DNA replication. P4 also controls replication when it propagates in the plasmid state. This control is achieved (i) by modulating the expression of the α gene at the transcriptional level (see below) and (ii) by the negative control on P4 DNA replication exerted by the product of the P4 *cnr* (copy number regulation) gene (101). The *cnr* gene is located immediately upstream of the α gene within the same operon and the expression of the two genes is coregulated (figure 26-3A) (31, 79). Using a two-hybrid system in yeast it was found that Cnr and gpα interact with each other (104). Overproduction of the Cnr protein inhibits P4 DNA replication, whereas lack of Cnr causes overreplication and host cell killing (101). Overall, it appears that a balanced expression of the two proteins is necessary for proper P4 DNA replication (101, 104, 110). Although cell death does not impair the P4 lytic cycle, a *cnr* defective P4 cannot be maintained as a plasmid. The *cnr* gene is therefore the only P4 gene essential exclusively for plasmid propagation.

P4 mutants insensitive to negative Cnr control (α-cr mutants) have been isolated. All such mutants carry amino acid substitutions in the C-terminal region of gpα (figure 26-4A) (110) and map to the dimerization interface of the DNA-binding domain (107). No interaction between α-cr mutant proteins and the Cnr protein could be detected in the two-hybrid system (104). These observations indicate that the α protein is the target of Cnr-mediated negative regulation and suggest that Cnr may act by interfering with dimerization of the gpα origin binding domain.

It was shown in vitro that the Cnr protein does not bind DNA but stimulates the binding affinity of wild-type α protein to *ori1* and *crr*. The α-cr mutant proteins are still able to bind specifically to *ori1* and *crr*, but in the presence of Cnr the binding of α-cr protein to *ori1* and *crr* is less stimulated than the binding of wild-type α protein. No apparent effect of Cnr on α primase and helicase activities was found (110). In addition, Cnr protein interferes with α–α interaction (104, 107), suggesting that the Cnr–α–DNA complex is not competent for replication initiation unless Cnr is released from the complex. The negative regulation might interfere with some essential steps, for example α–α interaction and/or *ori1* melting. A model for the control of P4 replication initiation by Cnr is presented in figure 26-5.

Two Replicons Coexist in P4

Deletion of *ori1* and the *cnr* gene from a minimal *oriI* replicon revealed the presence of an alternative replicon, *oriII* (102). Replication of *oriII* depends upon α protein, requires two *cis*-acting sites, *crr* and *ori2*, and, unlike *oriI*, is completely inhibited by Cnr protein. The *ori2* site has been mapped within a 22 bp region, internal to the α-gene coding sequence, approximately at the boundary between the primase and helicase domains (103) (I. Oliva and D. Ghisotti, unpublished data) (figure 26-4C). This observation suggests that the multifunctional α gene might originate from the fusion of two ancestral genes, coding for primase and helicase-DNA binding functions, respectively, separated by the origin of replication.

The essential elements composing the two mini-replicons are diagrammed in figure 26-4B. Both α helicase and primase functions are required for replication from *oriII* (103). The *ori2* site does not contain type I iterons and does not bind α protein (103). Moreover, no putative binding sites for other known P4 or *E. coli* factors, such as a DnaA-box

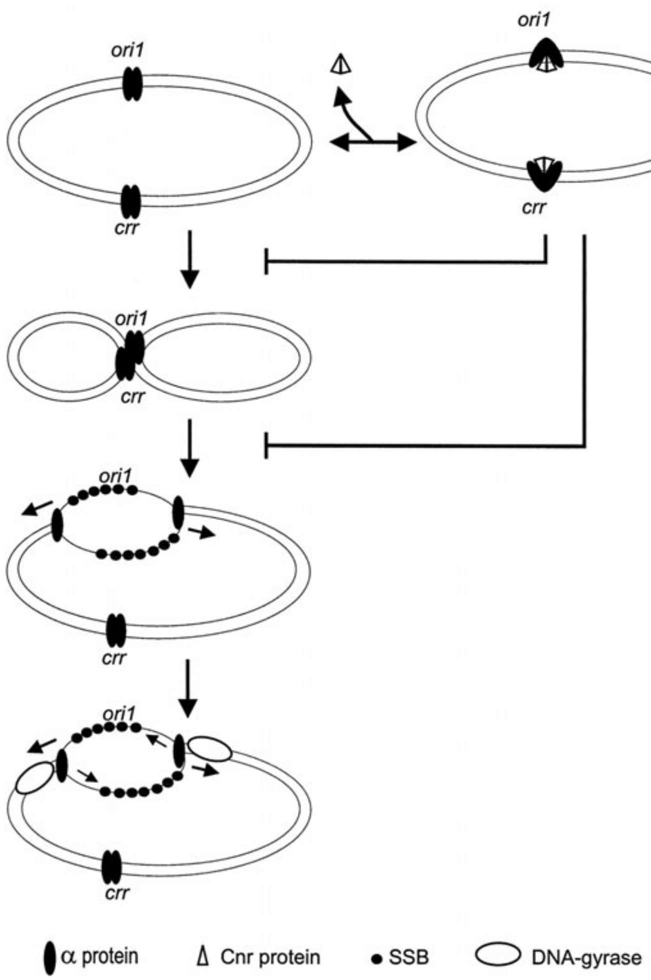


Figure 26-5 A model for P4 replication. Protein α binds to both *ori1* and *crr* sites in oligomeric form and may cause looping of P4 DNA by protein–protein interactions. Protein α is thought to make RNA primers (arrowheads) both at the origin of replication (priming) and at the replicative forks for lagging-strand priming (elongation). The Cnr protein, interacting with α, prevents replication initiation by inhibiting either DNA looping or origin denaturation and priming.

consensus sequence, are found. Thus, *ori2* differs structurally and functionally from *ori1*. This suggests that *oriI* and *oriII* replicons may replicate by different mechanisms.

The initiation site of replication in the *oriII* origin is still unidentified. In experiments in vitro aimed at mapping the replication start point in the P4 genome, a replication initiation signal was identified immediately to the right of *ori2* (63). This signal might represent the replication initiation point of the *oriII* replicon.

Several prokaryotic chromosomes (e.g., plasmid R6K and phage T7) contain multiple origins of DNA replication (25, 53, 100). Usually the major origins are used more than 90% of the time, while minor origins are either used infrequently or they remain silent. In the wild-type P4, *cnr*, coexpressed with α, is inhibitory on *oriII*, suggesting that this origin is normally silenced (102). Whether the *oriII* replicon is a relic of a P4 ancestor or whether it is still functional under some physiological conditions remains to be ascertained.

The Regulatory Network

The different modes of P4 propagation require the differential expression of replication and morphogenetic functions. This is achieved mainly through complex regulation of the α operon, which encodes both immunity and replication functions in the 5′ and 3′ regions, respectively, and the “late” *sid* operon. Three regulatory conditions have been described (see figure 26-3):

(i) The uncommitted phase, which immediately follows P4 infection of *E. coli*, is characterized by an early transcriptional burst of the entire α operon from the P_{LE} promoter, followed by the activation of a negative control system that causes premature transcription termination about 300 nucleotides downstream of P_{LE} (immunity control, see below). This regulatory mechanism is irreversible and marks the end of the uncommitted phase. One of two alternative states follows:

(ii) The immune-integrated (lysogenic) state, in which P4 gene expression is limited to the P_{LE} proximal part of the α operon (immunity region). Replication genes are therefore not expressed and the P4 integrated genome is passively replicated by the bacterial chromosome.

(iii) The lytic-plasmid state, in which the positively regulated late promoters P_{LL} and P_{sid} are activated by positive regulators. In this state, transcription from P_{LL} drives a new mode of expression of the entire α operon.

The key aspects of the regulatory circuitry that controls P4 gene expression in its different developmental phases are: (i) The α operon is downstream of the strong constitutive promoter P_{LE}. As a consequence, expression of the replication genes from P_{LE} is regulated at a post-transcription initiation level by controlling elongation and termination. (ii) A highly specific mechanism that efficiently terminates all transcripts from P_{LE} is ineluctably and irreversibly activated soon after infection. Thus, subsequent activation of replication genes, either in the late lytic phase and in the plasmid state or upon induction of P4 prophage by the helper, requires this mechanism to be circumvented, rather than reversed. This occurs by activating the positively regulated P_{LL} promoter. (iii) Because P_{LL} is upstream of P_{LE}, expression of the α operon from P_{LL} requires the termination roadblock imposed by immunity to be bypassed. This is achieved by translational suppression of transcription termination. (iv) Regulatory cross-talk permits P4 to sense the presence of P2 and to adjust accordingly both satellite and helper gene expression.

The Uncommitted Phase and the Path to Lysogeny

In the early, uncommitted phase the α operon is transcribed from the constitutive promoter P_{LE}, the replication genes are expressed at high level, and a burst of P4 replication is observed. Transcription from P_{LE} appears to be termination-prone and yields RNA molecules of different lengths (see figure 26-3A, B): (i) full-length mRNA, which covers the entire operon, (ii) transcripts that stop at t_{151}, a terminator just upstream of the replication genes *cnr* and α, and (iii) a family of transcripts less than 0.5 kb long which do not extend beyond two termination sites, t_4 and t_{imm}, that are located within the *kil* gene (13, 14, 31). Thus, different portions of the α operon are expressed differentially.

About 15 minutes after infection, transcripts extending beyond t_4–t_{imm} can no longer be detected and all transcripts from P_{LE} terminate either at t_1, an intrinsic terminator located about 70 bp downstream of P_{LE}, or at t_4–t_{imm}. In this way transcription from P_{LE} allows the expression of the replication genes for a restricted time after P4 infection, after which transcription is limited to the 5′ untranslated portion of the operon which encodes the immunity functions (immunity region, figure 26-3B). At this point the immune mode of transcription from P_{LE} is irreversibly established.

Transition from the early transcription pattern, with expression of the entire α operon, to the immune pattern is concomitant with the appearance of a small stable RNA 79 ±1 nucleotides long, the CI RNA, encoded by the *cI* gene (figures. 26.3B, 26.6A). The CI RNA is produced by processing of transcripts that cover the immunity region. Moreover, the presence in the cell of the CI RNA is sufficient to efficiently cause premature transcription termination at t_1 and t_4–t_{imm} in an infecting phage: thus the CI RNA is the P4 immunity factor (13, 15, 31, 37, 39, 42).

Immunity

For most known temperate bacteriophages, immunity is elicited by a repressor protein which prevents transcription initiation at promoter(s) controlling the expression of lytic functions. The P4 immunity mechanism is unique among the known bacteriophages in several respects: (i) expression of the P4 α operon is prevented by premature termination of transcription starting at the constitutive promoter, P_{LE}; (ii) the P4 immunity factor is a short, stable RNA (CI RNA); (iii) transcription termination is controlled via RNA–RNA interactions between the CI RNA and two specific target sequences on the nascent transcript; (iv) the CI immunity factor is produced by specific processing of the same transcript it controls.

Most recessive mutations that cause an immunity defect map in *seqB*, a region internal to the *cI* gene. *seqB* exhibits complementarity with two sequences, *seqA* and *seqC*, located upstream and downstream of *seqB*, respectively (figure 26-3B) (42, 85). The complementarity between *seqB* and both *seqA* and *seqC* must be maintained in order to achieve timely and efficient transcription termination. Moreover, mutations in both *seqA* and *seqC* that restore complementarity with a *cI* mutation in *seqB* also re-establish efficient transcription termination (42, 85). As a result, the *seqA* and *seqC* sites in the nascent transcript appear to be the targets of the CI RNA, as both are required for the establishment and the maintenance of prophage immunity. Interactions between the CI RNA and target RNAs appear to occur primarily between complementary regions that computer analysis predicts to be single-stranded (14, 85) (figure 26-6C).

The above model is further supported by the analysis of the immunity system of phage ΦR73. This P4-like phage is heteroimmune to P4 and has provided the opportunity

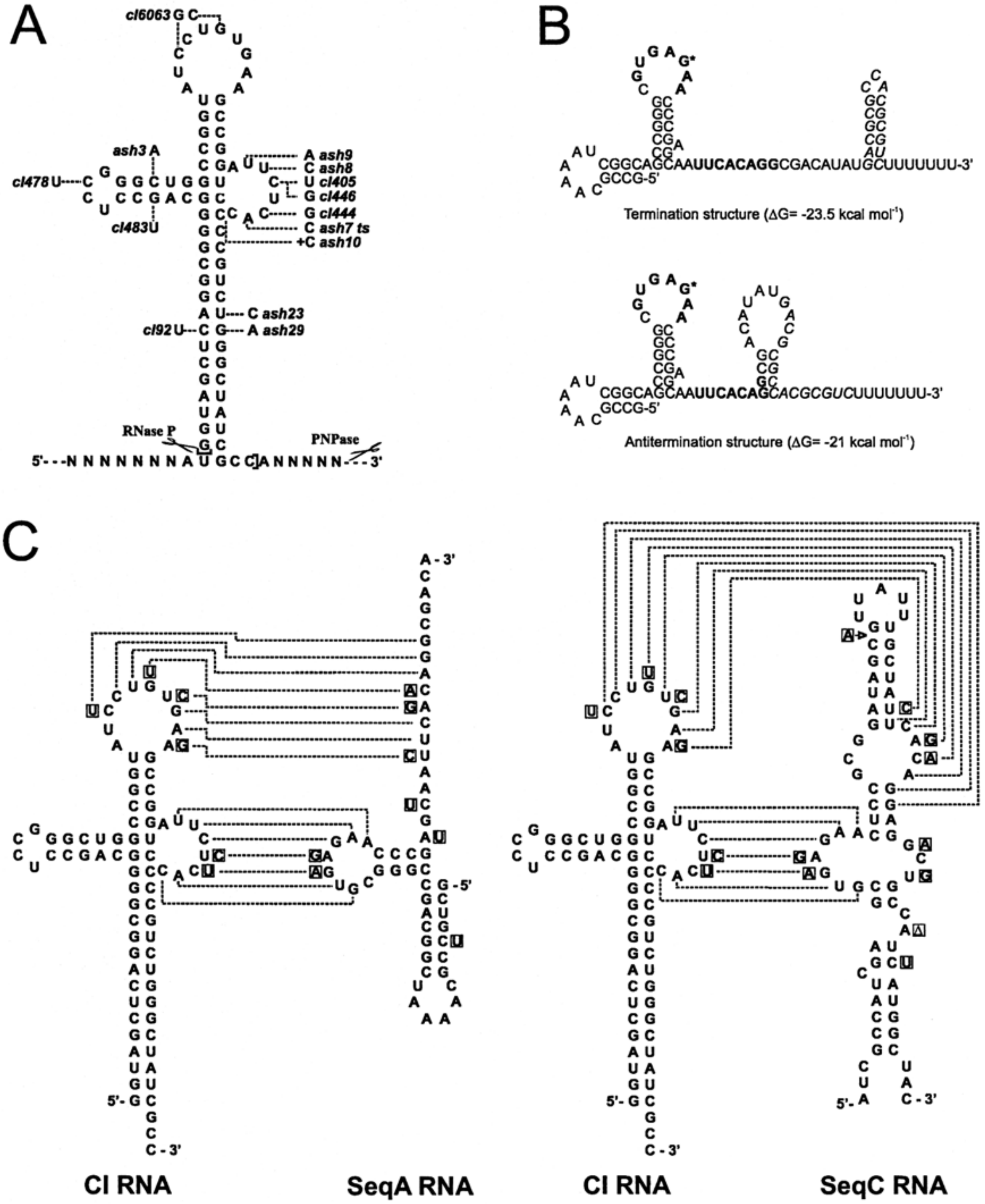

A
cl6063
ash3
cl478
cl483
A ash9
C ash8
U cl405
G cl446
G cl444
C ash7 ts
+C ash10
C ash23
cl92
A ash29
RNase P
PNPase
5'---NNNNNNNNAUGCC]ANNNNNN---3'
B
CGGCAGCAAUUCACAGGCGACAUAUGCUUUUUUU-3'
GCCG-5'
Termination structure (ΔG= -23.5 kcal mol-1)
CGGCAGCAAUUCACAGCACGCGUCUUUUUUU-3'
GCCG-5'
Antitermination structure (ΔG= -21 kcal mol-1)
C
A - 3'
5'- G C
C - 3'
C G - 5'
5'- A
C - 3'
CI RNA
SeqA RNA
CI RNA
SeqC RNA

Figure 26-6 Predicted secondary structure of the CI RNA and possible interactions in the leader transcript of the immunity region. A: Computer predicted structure of the CI RNA. The minimal sequence necessary for CI RNA maturation is reported and the 5′- and 3′ ends of the mature CI RNA molecule are indicated by brackets. Mutations that affect CI activity or processing are indicated; *cl478* is a silent mutation and *cl483* is a suppressor of *ash3* (39). From (39), with permission. B: Predicted secondary structures of the P_{LE}–t_1 transcript. The alternative structures and their free energy are reported. Boldface: bases complementary to CI RNA single-stranded regions; italics: bases involved in t_1 stem formation. From (14), with permission. C: Possible interactions between the CI RNA and the SeqA and SeqC targets in the leader region. The complementary bases in single stranded regions of CI RNA are connected by lines. The base changes in the P4-like retronphage ΦR73 are boxed. The arrow indicates an insertion, △ a deletion. Redrawn from (86), with permission.

to compare similar immunity systems with altered specificity (86). The ΦR73 CI RNA differs from the P4 CI RNA in six bases, all in *seqB*: two in the bulge and four in the major loop. It is noteworthy that for five of these changes there is a complementary base substitution in both *seqA* and *seqC*, whereas the sixth, a C-to-U transition in the major loop, is still compatible with pairing with the G in *seqA* and *seqC* (figure 26-6C).

How these RNA–RNA interactions between *seqB* and its targets, *seqA* and *seqC*, control transcription elongation is not completely understood. The complementary sequences in the immunity region may allow both intra- and intermolecular pairing of RNA molecules. In the uncommitted phase, transcription of the entire α operon occurs before the appearance of CI RNA. As soon as the mature CI RNA is produced in the cell, the efficiency of termination within the immunity region increases, completely preventing transcription of the downstream replication genes (13, 14, 31). These findings suggest that intramolecular interactions in the leader sequence allow read-through, whereas intermolecular interactions between the CI RNA and the nascent leader transcript cause strong termination.

seqA is located immediately upstream of t_1. It has been shown that termination at t_1 is enhanced by the CI RNA both in vivo and in vitro; CI-dependent termination at t_1 requires a wild-type *seqA* sequence, strongly suggesting that the pairing between *seqA* and the CI RNA may induce termination at t_1. It has been proposed that this interaction might prevent the formation of an antitermination structure on the nascent transcript, favoring t_1 folding (figure 26-6B). As t_1 is located upstream of *cI*, the CI RNA can autoregulate its own expression by reducing transcription of the *cI* gene (14).

seqC overlaps the ribosome binding site and the start codon of *kil*, the first translated gene in the P_{LE} transcript. When the nascent transcript is translated, termination at t_4 and t_{imm} does not occur. This suggests that *seqB–seqC* could indirectly control termination at these sites by regulating *kil* translation. In particular, lack of translation may allow the Rho factor access to t_{imm}, thus causing transcription termination by polar effect (38). t_1 and t_{imm} are not essential for the maintenance of immunity, whereas t_4 appears to be the main terminator necessary to prevent expression of the lytic genes in the lysogenic condition. The molecular mechanism of termination at t_4, which does not show any feature of either Rho-dependent or intrinsic terminator, remains to be clarified (14).

CI RNA maturation from the primary P_{LE} transcript requires RNase P and polynucleotide phosphorylase (PNPase). RNase P, an endonucleolytic ribozyme also involved in maturation of tRNAs, generates the 5′ end of the mature molecule (37). The 3′ end maturation is promoted by PNPase, and seems to be facilitated by polyadenylpolymerase I (PAP I). Interestingly, mutants in PNPase and PAP I, which are involved in 3′ end processing, are also defective in 5′ end RNA maturation of CI RNA, suggesting that 3′ end processing is required to generate the substrate for RNase P. RNase E also participates in the maturation of CI RNA, although its role is not essential (15, 76).

It is interesting to note that a similar regulatory module, based on transcription termination controlled by a small pseudo-antisense RNA, is structurally and functionally conserved in other systems unrelated to P4, such as bacteriophage N15 and P1, which also share with P4 the capability to be maintained as autonomously replicating plasmids. In all these systems the genes immediately downstream of and controlled by the CI-like RNAs are functionally related: the first gene is an inhibitor of cell division (*kil* in phages P4 and N15, *icd* in phage P1) and the second is an anti-repressor protein (in P4 the ε gene encodes the anti-repressor of the P2 helper phage, whereas P1 and N15 express an anti-self-immunity system (23, 40, 82, 84)). Functional and structural similarity, however, is not accompanied by sequence similarity, which is limited to the CI-like RNA. We think that these operons derive from an ancestral “anti-immunity module” that was acquired through horizontal gene transfer and evolved independently under different selective pressure, and that the regulatory RNA, which has been conserved to a greater extent, was subject to more stringent structural and functional constraints.

The control of gene expression found in these “anti-immunity modules,” based on a “pseudo-antisense” RNA that controls transcription termination, appears to be very effective in obtaining a transient expression of potentially lethal genes upon infection of a naïve cell. By sequence-structure similarity searches in sequence databases, potential CI-like RNAs have been found in a *Shigella flexneri* prophage-related sequence, in IncI1 plasmids, and in the *Acinetobacter* chromosome (38, 82), suggesting a widespread diffusion of this mechanism of genetic control.

Integration

In addition to establishing the immune condition that prevents expression of the replication genes, lysogenization by P4 requires the integration of the P4 genome into the bacterial chromosome. Integration occurs according to the Campbell model by site-specific recombination between the P4 and the bacterial *att* sites (see chapter 7); the latter corresponds to the 3′ end of a gene (*leuX*) encoding a $tRNA^{Leu}$ isoacceptor. Recombination occurs within a 20 nucleotide long, GC-rich core region identical in both phage and host *att* sites, and this preserves the integrity of the *leuX* gene sequence upon prophage integration (78).

Integration requires the integrase, coded by the P4 *int* gene (17a, 77) (figure 26-3A). This gene, with coding capacity for a 426 amino acid protein, is located to the left of *attP* and is transcribed leftwards (78) (D. Piazzolla and D. Ghisotti, unpublished data). P4 integrase belongs to the highly divergent family of site-specific recombinases, which includes the well-characterized λ integrase (3). The C-terminal half of the protein, which is directly involved in the recombination event, is particularly well conserved in all integrases of the λ family, whereas the N-terminal region that specifically binds the arm sequences of the *att* sites is more divergent. In the P4 *att* site a pair of 16 bp direct repeats, present on either side of the core sequence, and an inverted repeat in the left arm are supposed to be bound by Int (78). A consensus sequence for IHF is present in the right arm of *attP*, suggesting that this bacterial protein is part of the P4 integration complex.

The *int* promoter shows canonical σ^{70} consensus sequences and is active early after P4 infection. At later stages and in the lysogenic state, P_{int} activity is low (17b); it has been suggested that Int autoregulates its own expression by binding to the *attP* left arm sequences that overlap the P_{int} region (78).

P4 integrase is necessary for both P4 integration and excision (77, 78). This latter event occurs spontaneously at low frequency and, at much higher efficiency, upon infection of a P4 lysogen by the helper phage P2 (87). Excision requires a second P4 protein, Vis, a small (88 amino acids) basic protein that presents a helix–turn–helix motif and binds the P4 *attP* region (17b, 79). Vis is expressed from P_{LL}, the P4 late promoter that is *trans*-activated by phage P2 (see below). Therefore, the first response of a P4 prophage to P2 infection is to promote excision and the lytic cycle.

Like most excisionase proteins, Vis binding causes bending of DNA (17b), suggesting that Vis actively participates in the formation of the protein–DNA complex involved in excision. Moreover, Vis is a regulatory protein that controls transcription from P_{LL}, P_{sid} (see below), and probably P_{int}, thus being involved in the control of *int* expression (79) (D. Piazzolla and D. Ghisotti, unpublished data).

Integration within known or putative tRNA genes appears to be a widespread phenomenon among both prokaryotic and eukaryotic mobile genetic elements such as viruses, plasmids, and transposons (19, 83); this may reflect the structural and/or mechanistic features offered by tRNA genes that may be exploited for evolution of integration systems.

P4-like cryptic prophages, or simply P4 integrase homologs integrated into a tRNA gene and associated with clusters containing accessory bacterial functions, have been found repeatedly in Gram-negative species. The functions associated with the above elements include retrons, modification-restriction systems, pathogenicity and symbiosis islands, and catabolic pathways (5, 9, 16, 18, 24, 49, 60, 61, 78, 81, 97, 99). These findings testify to a wide diffusion of P4-like genetic elements and their direct involvement in bacterial evolution.

The Late-Plasmid Transcription Pattern

In P4, the choice between lytic plasmid versus lysogenic development relies on the presence or absence of activation of the two late promoters, P_{LL} and P_{sid}. The lytic plasmid and the immune modes of transcription are not mutually exclusive; rather, the former modes are superimposed on the immunity control. Although the patterns of gene expression in the plasmid and lytic cycles are essentially superimposable, the mechanisms leading to the activation as well as the outcome of these two conditions are markedly different.

The Multicopy Plasmid State

Maintenance of P4 in the plasmid state requires that two main conditions are satisfied: (i) the immune regulation that prevents expression of the plasmid replication functions must be bypassed, and (ii) P4 plasmid replication must be controlled in order to maintain a number of P4 genomes compatible with survival of the bacterial host. In the plasmid state P_{LL} drives the expression of the α operon and thus its regulation is crucial for the establishment and maintenance of the plasmid condition. Homeostatic control of P_{LL} seems to be achieved by the opposing actions of gpδ, expressed from P_{sid}, and Vis, encoded by the first gene transcribed from P_{LL} itself. These two proteins activate and repress P_{LL}, respectively (29, 30, 79).

P_{LL} is located 400 nucleotides upstream of P_{LE} (30). When transcription starts at P_{LL} two additional genes, *vis* and *eta*, are expressed. Consequently, the transcription termination barrier imposed by the immunity system when transcription initiates at P_{LE} rather than P_{LL} is bypassed via translational suppression of transcription termination (38) (figure 26-3). The start codon of *vis* is located about 50 nucleotides downstream of P_{LL}; the start codon of *eta* is partially overlapped by the *vis* stop codon, and the two genes appear to be translationally coupled (38, 79). *eta* extends through the immunity region, overlapping *kil* in frame (Kil is therefore a truncated form of Eta). Thus, transcripts starting at P_{LL}, unlike transcripts starting at P_{LE}, do not have an untranslated leader region. Nonsense mutations in *vis* or *eta* cause premature termination of transcription about 700 nucleotides downstream of P_{LL}, suggesting that ribosome protection of the RNA transcribed from P_{LL} may prevent interaction between the transcript and the CI RNA and/or the Rho factor, thus impeding transcription termination (38). Therefore it appears that immunity is the default program inescapably expressed by an infecting P4 phage. To enter the alternative life-style (plasmid state or lytic cycle), P4 simply bypasses the immunity mechanism by activating termination-insensitive transcription from an upstream promoter, P_{LL}.

The termination insensitivity necessary to enter into the plasmid state does not extend to t_{151}, which is found downstream of the promoter P_{LE} (see figure 26-3). As a result, the expression of the α gene, which is found downstream of t_{151}, may still be inhibited. Differential expression of replication and morphogenetic functions, as controlled by the α protein, can consequently occur even while P4 is in the plasmid state (30, 38).

Upon infection by wild-type P4, the multicopy plasmid state may be established at a low frequency, the lysogenic cycle being preferred in the absence of a helper phage genome. Mutations affecting P4 immunity may increase the frequency of plasmid establishment. However, the inability to establish immunity does not necessarily direct P4 to plasmid propagation, since uncontrolled expression of the *kil* gene from P_{LE} is lethal to the host (38). As a consequence, P4 *cI* and P4 *seqA/seqC* mutants, in addition to being unable to lysogenize, kill the majority of the infected cells and enter the plasmid state at a low frequency, comparable to that of the wild-type phage (2, 85). On the contrary, the P4 *cI kil1* double mutant establishes the plasmid state in about 100% of infected cells (2). This would indicate that establishment of plasmid propagation requires the shut-off of *kil* expression.

Transcription from P_{LL} is essential for expression of plasmid replication functions in the plasmid condition. In fact, P4 *vis* nonsense mutants, in which transcription from P_{LL} terminates prematurely, and P4 δ mutants, in which P_{LL} cannot be activated, are impeded in plasmid propagation (29, 30, 38) whereas activation of transcription from P_{LL} by promoter-up mutations, which make transcription independent of the positive regulators, increase the frequency of P4 plasmid establishment (28, 30, 85).

The expression level of the plasmid replication genes from P_{LL} is lower than from P_{LE}. This appears to depend on negative regulation of P_{LL} transcription by the Vis protein: expression of a cloned Vis gene causes a dramatic inhibition of transcription from P_{LL}, it inhibits P4 propagation in the plasmid state, and suppresses the virulent phenotype of P4 *vir1*, a promoter-up mutation that makes P_{LL} independent of its activators (66, 79). The indirect control of the P4 plasmid replication rate due to negative regulation on P_{LL} transcription is not sufficient for P4 maintenance in the plasmid state. A more direct control on P4 plasmid replication due to the Cnr protein (see above) is necessary for copy number control of plasmid P4.

A low basal level of transcription from P_{LL} may be detected soon after infection, due to an overlapping weak promoter $P_{LL}{}^*$ (29). However, no basal transcription has been detected from P_{sid}. This poses the unsolved problem of how P_{sid} transcription (and, as a consequence, the plasmid regulatory state) is primed upon infection, when gpδ is not yet present in the cell. It is possible that the right operon may be transcribed at a very low basal level in a δ-independent manner. Alternatively, either the P4 Vis protein or other unidentified P4 functions might be involved.

A possible question is whether the P4 plasmid condition is simply an abortive lytic cycle of no biological significance. In the laboratory, in the absence of the helper, the lysogenic condition is largely preferred and is more stable, thanks to the efficient integration and immunity mechanisms, and maintaining the wild-type P4 as a plasmid requires the continuous selection of plasmid carrier clones, which grow more slowly than the lysogens. It therefore seems likely that also in nature lysogeny is the preferred condition and the plasmid state might be transiently established upon P4 infection of a host lacking a helper, or upon abortive infection of a P4 lysogen by a helper phage.

The Lytic Cycle

Satellite–Helper Regulatory Interactions

The P4 lytic cycle requires efficient exploitation of the helper phage genetic information to obtain the morphogenetic and lysis gene products. This exploitation may occur under different scenarios: (i) P4 infecting a repressed P2 lysogen; (ii) P4 and P2 co-infecting a non-lysogenic host; (iii) a P4 lysogen being infected by phage P2; (iv) a P4-plasmid carrier being infected by phage P2. In each of these situations, the satellite phage senses the presence of the helper and responds by activating its own functions that in turn will modify the pattern of gene expression of the helper. Reciprocal regulatory interactions developed in this satellite–helper system involve lifting the immunity mechanisms (mutual derepression) and direct activation of the late operons (reciprocal *trans*-activation) of both phages (see figure 26-7).

P2 prophage derepression by phage P4 occurs via direct inactivation of the helper repressor by the P4 ε gene product. It appears that the gp ε dimeric protein interacts with the dimeric C repressor of phage P2 to form gp ε–C multimeric complexes that interfere with the binding of P2 repressor to its operator (40, 70, 71, 93). On the contrary, P2 may derepress the P4 satellite prophage via activating the P_{LL} late promoter by the helper early gene product Cox, thus bypassing the P4 immunity system. The Cox binding region is located −60 to −150 in P_{LL} (28, 87, 93).

The P4 late promoters P_{LL} and P_{sid}, and the P2 late promoters of the morphopoietic operons, may be reciprocally *trans*-activated by the P4-encoded gpδ and the P2-encoded Ogr transcriptional activators (26, 30, 46). Ogr is the prototype of a class of transcriptional activators that control the late operons of phages related to either P2 or P4 (46). These activators are small proteins (72–81 amino acids) with a zinc-finger-like metal-binding domain. Like most members of this family, phage ΦR73 gpδ is composed of a single module (52, 56). Interestingly, P4 gpδ is twice

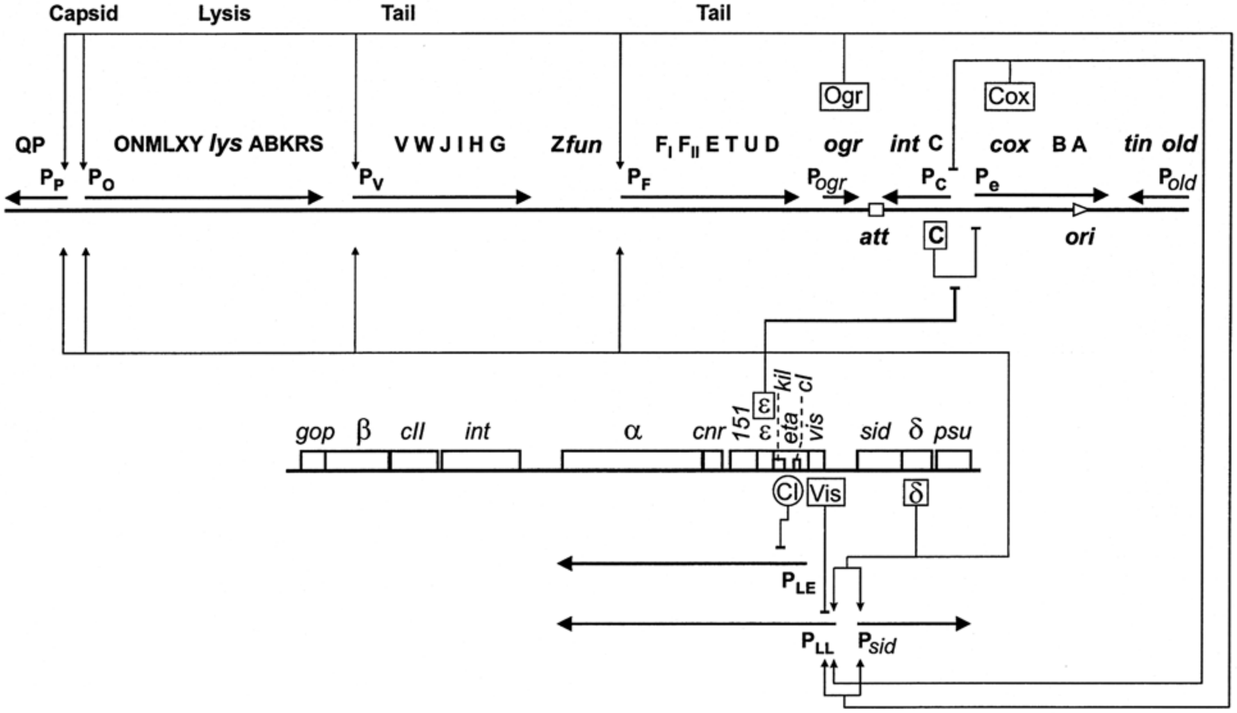


Figure 26-7 Regulatory loops controlling expression of P4 and P2 late genes. Straight arrows indicate transcription units. Transcriptional or functional activation and inhibition are indicated by bent lines ending with an arrowhead or with a minus sign, respectively. The P2 map is not drawn to scale. Redrawn from (69), with permission.

the size of the other members of the family; it contains two zinc-binding motifs and appears to be a tandem duplication of the basic Ogr module (48). Both domains of gpδ are required for transcriptional activity (57).

The Ogr-family activators interact with the C-terminal domain of the RNA polymerase α subunit and bind to a consensus sequence centered at about −55 nucleotides from the transcription start point in all the responsive σ^{70} -dependent promoters (45, 56). The P4 P_{sid} promoter, however, contains a second copy of the consensus sequence in the −18 region that appears to quench transcription activation by δ, as mutations in the P_{sid} −18 region increase δ-dependent promoter activity (83). This seems to explain why δ-promoted transcription is less efficient from P_{sid} than from P2 late promoters and might be instrumental in regulating δ expression in the plasmid condition.

In addition to these regulatory interactions, morphogenetic control is elicited by P4 in order to assemble a capsid of the correct size (see below). The type and timing of the satellite–helper crosstalk vary under the different infection conditions and give different outputs, in terms of the satellite versus helper virions produced, that appear to fit the reproductive strategies of the satellite phage. Details associated with these different infection conditions are considered individually as follows:

(i) *Infection of a P2 lysogen by P4.* When P4 infects a repressed P2 lysogen and the lytic option is set, the helper prophage immunity is lifted by P4 gpε. Phage P2 derepression leads to expression of the helper early genes (*cox* and the replication genes A and B), and to activation of unidirectional P2 DNA replication in situ, without excision of the integrated prophage genome (40, 93). P2 late gene expression (morphogenesis and lysis) may therefore ensue via the normal P2 activation mechanism, which requires P2 DNA replication and Ogr transcriptional activator. Expression of P2 morphogenesis and lysis genes is both necessary and sufficient for the completion of the P4 lytic cycle (40, 46, 95). P4, however, may efficiently *trans*-activate P2 late gene expression, bypassing the need for P2 replication and the positive regulator, Ogr. *Trans*-activation can occur because P4 gpδ activates P2 transcription from the same promoters as those used by the P2 phage Ogr transcriptional activator (21, 22, 46, 95).

The P4 *ε* gene is essential for P4 growth in a repressed P2 lysogen, but not in co-infection with P2 of a nonlysogenic cell (32). Derepression appears to be necessary for the timely activation of the helper's morphogenetic operons by the P2 Ogr and/or the P4 gpδ proteins. Either transcriptional activator is sufficient to *trans*-activate both P4 and P2 late genes, although in the presence of functional Ogr and gpδ proteins the P4 lytic cycle occurs more efficiently than with a single activator (95).

The low P2 excision frequency and the interference with P2 growth due to the efficient production of small capsids (7, 32, 89) together appear to be responsible for the production of mostly P4 rather than P2 virions. Thus, in an environment of P2 lysogenic bacteria, where P2 replication would be inhibited by the P2 prophage immunity, all the morphogenetic potential is efficiently directed toward P4 propagation.

(ii) *P4 and P2 co-infecting a non-lysogenic host.* When P4 and P2 co-infect a non-lysogenic host, the burst size of P2 is reduced and progeny of both phages is produced in about the same proportions. Interference with P2 growth is stronger if P4 infects the host earlier than P2. The gpε derepression activity is dispensable for the P4 lytic cycle, although it appears to be required to interfere with the helper's growth (32).

(iii) Infection of P4 lysogen by P2. Infection of a P4 lysogenic strain by phage P2 virions induces the P4 lytic cycle. However, the yield of the P4 phage is low (ca. 1 per cell) and interference on P2 growth is not observed. The P4 genome can therefore be rescued under such conditions, albeit at low efficiency, and can survive phage P2-mediated host cell death. The P4 prophage senses the infection of the helper via the P2 early protein, Cox, which activates P_{LL} and thus the expression of the α operon. Cox does not activate P_{sid}, which lacks the Cox binding site. At a later stage after infection, P_{sid} may be efficiently activated by the P2 Ogr protein (87, 93).

(iv) *Infection of a P4-plasmid carrier by P2.* When P2 infects a cell carrying P4 in the plasmid condition, the Cox function is not required to induce P4 lytic cycle since the P4 late promoters are already activated (2, 28). In such conditions P4 strongly interferes with P2 growth.

P4 Morphogenesis

The final step in satellite–helper interactions is at the level of viral particle assembly. Phage P2 packages its 33 kb genome into 60 nm isometric icosahedral capsids consisting of 60 hexameric and 12 pentameric capsomers (T = 7). The capsomers are made of the N-terminal processed products (gpN*) of the P2 *N* gene. The capsid maturation pathway involves (i) assembly of a procapsid made of gpN, the internal scaffolding protein gpO, and the tail connector at one vertex; (ii) proteolytic processing of the procapsid proteins accompanied by the structural reorganization of the procapsid shell lattice; (iii) packaging of the 33 kb long P2 DNA molecule. Pre-assembled tails are then added to the connector vertex (see chapter 25). P4 subverts the P2 morphogenetic pathway by directing the assembly of a 45 nm isometric capsid (consisting of 30 hexamers and 12 pentamers, T = 4) more suitable for its own genome, which is a third of the phage P2 genome size. A schematic model for P4 and P2 capsid assembly is shown in figure 26-1C.

Assembly of Small Heads

The choice of the T = 4 morphogenetic pathway is dictated by the presence of the P4 *sid* (size determination) gene product. P4 *sid* mutants are unable to form small capsids. They package dimers and trimers of P4 DNA into P2-size heads (89). Mutations in P2 that make the head assembly process resistant to the effect of Sid map in P2 *N* gene (*sir* mutants: sid responsiveness) (94). The P2 Sir phenotype may be suppressed by second-site mutations in P4 *sid* (supersid mutations; (59). This genetic analysis suggests that P4 Sid protein interacts with P2 gpN and alters the "normal" P2-directed head assembly geometry. Coexpression in *E. coli* of phage P2 *N* and P4 *sid* genes only, leads to the production of small procapsid-like particles (35, 73, 74). In three-dimensional reconstructions, Sid forms a continuous dodecahedral scaffold on the outside of the particles. This external scaffold connects at the 3-fold axes and makes contacts with the underlying shell on the gpN hexamers (74) (figure 26-1B). Similar structures may be obtained in vitro from purified gpN and Sid proteins and in vivo upon coexpression of both genes (35, 105). It is plausible that Sid acts on the subassembly of the hexameric capsomers which, in turn, determine the final geometry of hexamer and pentamer assembly. It should be noted, however, that in vivo the Sid external scaffold does not substitute for the P2-encoded gpO internal scaffolding protein, which is essential for the production of both P2 and P4 virions (65, 92).

Maturation of the procapsid involves the interdependent proteolysis of the P2 gpN shell and gpO internal scaffold proteins, and the release of the Sid external scaffold (65, 72, 73). In addition, the Psu capsid "decoration" protein associates on the top of the hexameric capsomers and stabilizes the phage particle. Psu is the only P4-encoded protein that may be found in the mature P4 virion, but it is not essential for P4 morphogenesis (54, 55).

DNA Packaging

P4 dependence on the helper for the morphogenetic pathway includes DNA maturation and packaging. Circular DNA molecules, the product of both P2 and P4 replication pathways, are the preferred substrate for packaging by both P2 and P4 proheads. A staggered double-strand cut that generates the 19 nucleotide long single-stranded cohesive ends of the mature genome occurs within a 55 nucleotide long sequence, essential for DNA packaging, that is identical in P2 and P4 (10, 80, 112) (see chapter 25 and references therein).

From the above description it is apparent that P4 has developed sophisticated regulatory strategies to exploit P2 and maximize the chances for its own horizontal transmission in bacterial populations that may differ with regard to the presence and the state of the helper phage. Clearly, P4 appears as an episome that can propagate horizontally by a specialized system of "generalized transduction." We speculate that P4 evolved from an ancestral plasmid replicon by independent acquisition of an "integration module" (*att*, *int*) for its stable maintenance in the host chromosome, and a "transduction module" (*cos* and, probably later on, *sid*) for packaging in the transducing helper phage. The interaction with the helper phage has been the driving force in the evolution of the lytic-plasmid regulatory circuitry that adapted to the helper-phage late regulatory mechanisms.

REFERENCES

1. Agarwal, M., M. Arthur, R. Arbeit, and R. Goldstein. 1990. Regulation of icosahedral virion capsid size by the in vivo activity of a cloned gene product. Proc. Natl. Acad. Sci. USA 87:2428–2432.
2. Alano, P., G. Dehò, G. Sironi, and S. Zangrossi. 1986. Regulation of the plasmid state of the genetic element P4. Mol. Gen. Genet. 203:445–450.
3. Argos, P., A. Landy, K. Abremski, J. B. Egan, E. Haggård-Ljungquist, R. H. Hoess, M. L. Kahn, B. Kalionis, S. V. Narayana, III, L. S. Pierson, N. Sternberg, and J. M. Leong. 1986. The integrase family of site-specific recombinases: regional similarities and global diversity. EMBO J. 5:433–440.
4. Arnold, H. P., Q. She, H. Phan, K. Stedman, D. Prangishvili, I. Holz, J. K. Kristjansson, R. Garrett, and W. Zillig. 1999. The genetic element pSSVx of the extremely thermophilic crenarchaeon *Sulfolobus* is a hybrid between a plasmid and a virus. Mol. Microbiol. 34:217–226.
5. Bach, S., C. Buchrieser, M. Prentice, A. Guiyoule, T. Msadek, and E. Carniel. 1999. The high-pathogenicity island of *Yersinia enterocolitica* Ye8081 undergoes low-frequency deletion but not precise excision, suggesting recent stabilization in the genome. Infect. Immun. 67:5091–5099.
6. Barrett, K. J., W. Gibbs, and R. Calendar. 1972. A transcribing activity induced by satellite phage P4. Proc. Natl. Acad. Sci. USA 69:2986–2990.
7. Bertani, L. E. 1968. Abortive induction of bacteriophage P2. Virology 36:87–103.
8. Bertani, G., and G. Dehò. 2001. Bacteriophage P2: recombination in the superinfection preprophage state and under replication control by phage P4. Mol. Gen. Genet. 266:406–416.
9. Blattner, F. R., G. Plunkett, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N.W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao. 1997. The complete genome sequence of *Escherichia coli* K-12. Science 277:1453–1474.
10. Bowden, D. W., R. S. Twersky, and R. Calendar. 1975. *Escherichia coli* deoxyribonucleic acid synthesis mutants: their effect upon bacteriophage P2 and satellite bacteriophage P4 deoxyribonucleic acid synthesis. J. Bacteriol. 124:167–175.
11. Bramhill, D., and A. Kornberg. 1988. A model for initiation at origins of DNA replication. Cell 54:915–918.
12. Brendler, T., A. Abeles, and S. Austin. 1991. Critical sequences in the core of the P1 plasmid replication origin. J. Bacteriol. 173:3935–3942.
13. Briani, F., S., Zangrossi, D. Ghisotti, and G. Dehò. 1996. A Rho-dependent transcription termination site regulated by bacteriophage P4 RNA immunity factor. Virology 223:57–67.
14. Briani, F., D. Ghisotti, and G. Dehò. 2000. Antisense RNA-dependent transcription termination sites that modulate lysogenic development of satellite phage P4. Mol. Microbiol. 36:1124–1134.
15. Briani, F., E. Del Vecchio, M. Migliorini, E. Hajnsdorf, P. Régnier, D. Ghisotti, and G. Dehò. 2002. RNase E and polyadenyl polymerase I are involved in maturation of CI RNA, the P4 phage immunity factor. J. Mol. Biol. 318:321–331.
16. Buchrieser, C., R. Brosch, S. Bach, A. Guiyoule, and E. Carniel. 1998. The high-pathogenicity island of *Yersinia pseudotuberculosis* can be inserted into any of the three chromosomal *asn* tRNA genes. Mol. Microbiol. 30:965–978.

17a. Calendar, R., E. Ljungquist, G. Dehò, D. C. Usher, R. Goldstein, P. Youderian, G. Sironi, and E. W. Six. 1981. Lysogenization by satellite phage P4. Virology 113:20–38.

17b. Cali, S., E. Spoldi, D. Piazzolla, I. B. Dodd, F. Forti, G. Dehò and D. Ghisotti. 2004. Bacteriophage P4 Vis protein is needed for prophage excision. Virology 322:82–92.

18. Cheetham, B. F., and M. E. Katz. 1995. A role for bacteriophages in the evolution and transfer of bacterial virulence determinants. Mol. Microbiol. 18:201–208.
19. Cheetham, B. F., D. B. Tattersall, G. A. Bloomfield, J. I. Rood, and M. E. Katz. 1995. Identification of a gene encoding a bacteriophage-related integrase in a *vap* region of the *Dichelobacter nodosus* genome. Gene 162:53–58.
20. Christian, R. B. 1990. Studies on P4 bacteriophage replication. Thesis, University of California, Berkeley.
21. Christie, G. E., and R. Calendar. 1983. Bacteriophage P2 late promoters. Transcription initiation sites for two late mRNAs. J. Mol. Biol. 167:773–790.
22. Christie, G. E., and R. Calendar, 1985. Bacteriophage P2 late promoters. II. Comparison of the four late promoter sequences. J. Mol. Biol. 181:373–382.
23. Citron, M., and H. Schuster. 1990. The c4 repressors of bacteriophages P1 and P7 are antisense RNAs. Cell 62:591–598.
24. Clark, A. J., J. Beltrame, and P. A. Manning. 1991. The *oac* gene encoding a lipopolysaccharide O-antigen acetylase maps adjacent to the integrase-encoding gene on the genome of *Shigella flexneri* bacteriophage Sf6. Gene 107:43–52.
25. Crosa, J. H. 1980. Three origins of replication are active in vivo in the R plasmid RSF1040. J. Biol. Chem. 255:11075–11077.
26. Dale, E. C., G. E. Christie, and R. Calendar. 1986. Organization and expression of the satellite bacteriophage P4 late gene cluster and the sequence of the polarity suppression gene. J. Mol. Biol. 192:793–803.

27. Dehò, G. 1983. Circular genetic map of satellite bacteriophage P4. Virology 126:267–278.
28. Dehò, G., D. Ghisotti, P. Alano, S., Zangrossi, M. G. Borrello, and G. Sironi. 1984. Plasmid mode of propagation of the genetic element P4. J. Mol. Biol. 178:191–207.
29. Dehò, G., D. Ghisotti, S. Zangrossi, P. Alano, T. Neri, S. Fattore, and G. Sironi. 1988. Regulation of alternative intracellular states in the phage-plasmid P4, pp. 65–72. *In* M. Bissel, G. Dehò, G. Sironi, and A. Torriani. (eds.) Gene Expression and Regulation: The Legacy of Luigi Gorini. Excerpta Medica, Amsterdam.
30. Dehò, G., S. Zangrossi, D. Ghisotti, and G. Sironi. 1988. Alternative promoters in the development of bacteriophage plasmid P4. J. Virol. 62:1697–1704.
31. Dehò, G., S. Zangrossi, P. Sabbattini, G. Sironi, and D. Ghisotti. 1992. Bacteriophage P4 immunity controlled by small RNAs via transcription termination. Mol. Microbiol. 6:3415–3425.
32. Diana, C., G. Dehò, J. Geisselsoder, L. Tinelli, and R. Goldstein. 1978. Viral interference at the level of capsid size determination by satellite phage P4. J. Mol. Biol. 126:433–445.
33. Diaz-Orejas, R., G. Ziegelin, R. Lurz, and E. Lanka. 1994. Phage P4 DNA replication in vitro. Nucleic Acids Res. 22:2065–2070.
34. Dokland, T. 1999. Scaffolding proteins and their role in viral assembly. Cell. Mol. Life Sci. 56:580–603.
35. Dokland, T., S. Wang, and B. H. Lindqvist. 2002. The structure of P4 procapsids produced by co-expression of capsid and external scaffolding proteins. Virology 298:224–231.
36. Flensburg, J., and R. Calendar. 1987. Bacteriophage P4 DNA replication. Nucleotide sequence of the P4 replication gene and the *cis* replication region. J. Mol. Biol. 195:439–445.
37. Forti, F., P. Sabbattini, G. Sironi, S. Zangrossi, G. Dehò, and D. Ghisotti. 1995. Immunity determinant of phage-plasmid P4 is a short processed RNA. J. Mol. Biol. 249:869–878.
38. Forti, F., S. Polo, K. B. Lane, E. W. Six, G. Sironi, G. Dehò, and D. Ghisotti. 1999. Translation of two nested genes in bacteriophage P4 controls immunity-specific transcription termination. J. Bacteriol. 181:5225–5233.
39. Forti, F., I. Dragoni, F. Briani, G. Dehò, and D. Ghisotti. 2002. Characterization of the small antisense CI RNA that regulates bacteriophage P4 immunity. J. Mol. Biol. 315:541–549.
40. Geisselsoder, J., P. Youderian, G. Dehò, M. Chidambaram, R. Goldstein, and E. Ljungquist. 1981. Mutants of satellite virus P4 that cannot derepress their bacteriophage P2 helper. J. Mol. Biol. 148:1–19.
41. Ghisotti, D., S. Finkel, C. Halling, G. Dehò, G. Sironi, and R. Calendar. 1990. Nonessential region of bacteriophage P4: DNA sequence, transcription, gene products, and functions. J. Virol. 64:24–36.
42. Ghisotti, D., R. Chiaramonte, F. Forti, S. Zangrossi, G. Sironi, and G. Dehò. 1992. Genetic analysis of the immunity region of phage-plasmid P4. Mol. Microbiol. 6:3405–3413.
43. Goldstein, R., J. Lengyel, G. Pruss, K. Barrett, R. Calendar, and E. Six. 1974. Head size determination and the morphogenesis of satellite phage P4. Curr. Top. Microbiol. Immunol. 68:59–75.
44. Goldstein, R., J. Sedivy, and E. Ljungquist. 1982. Propagation of satellite phage P4 as a plasmid. Proc. Natl. Acad. Sci. USA 79:515–519.
45. Grambow, N. J., N. K. Birkeland, D. L. Anders, and G. E. Christie. 1990. Deletion analysis of a bacteriophage P2 late promoter. Gene 95:9–15.
46. Halling, C., and R. Calendar. 1990. Bacteriophage P2 *ogr* and P4 *delta* genes act independently and are essential for P4 multiplication. J. Bacteriol. 172:3549–3558.
47. Halling, C., R. Calendar, G. E. Christie, E. C. Dale, G. Dehò, S. Finkel, J. Flensburg, D. Ghisotti, M. L. Kahn, and K. B. Lane. 1990. DNA sequence of satellite bacteriophage P4. Nucleic Acids Res. 18:1649.
48. Halling, C., M. G. Sunshine, K. B. Lane, E. W. Six, and R. Calendar. 1990. A mutation of the transactivation gene of satellite bacteriophage P4 that suppresses the *rpoA109* mutation of *Escherichia coli*. J. Bacteriol. 172:3541–3548.
49. Hare, J. M., A. K. Wagner, and K. A. McDonough. 1999. Independent acquisition and insertion into different chromosomal locations of the same pathogenicity island in *Yersinia pestis* and *Yersinia pseudotuberculosis*. Mol. Microbiol. 31:291–303.
50. Harris, J. D., and Calendar, R. 1978. Transcription map of satellite coliphage P4. Virology 85:343–358.
51. Inman, R. B., M. Schnös, L. D. Simon, E. W. Six, and D. H. J. Walker. 1971. Some morphological properties of P4 bacteriophage and P4 DNA. Virology 44:67–72.
52. Inouye, S., M. Sunshine, E. Six, and M. Inouye. 1991. Retronphage ΦR73: an *E. coli* phage that contains a retroelement and integrates into a tRNA gene. Science 252:969–971.
53. Inuzuka, N., M. Inuzuka, and D. R. Helinski. 1980. Activity in vitro of three replication origins of the antibiotic resistance plasmid RSF1040. J. Biol. Chem. 255:11071–11074.
54. Isaksen, M. L. S. T. Rishovd, R. Calendar, and B. H. Lindqvist. 1992. The polarity suppression factor of bacteriophage P4 is also a decoration protein of the P4 capsid. Virology 188:831–839.
55. Isaksen, M. L., T. Dokland, and B. H. Lindqvist. 1993. Characterization of the capsid associating activity of bacteriophage P4's Psu protein. Virology 194:674–681.
56. Julien, B., and R. Calendar. 1996. Bacteriophage PSP3 and ϕR73 activator proteins: analysis of promoter specificities. J. Bacteriol. 178:5668–5675.
57. Julien, B., P. Lefevre, and R. Calendar. 1997. The two P2 Ogr–like domains of the δ protein from bacteriophage P4 are required for activity. Virology 230:292–299.
58. Kahn, M., D. Ow, B. Sauer, A. Rabinowitz, and R. Calendar. 1980. Genetic analysis of bacteriophage P4 using P4-plasmid ColE1 hybrids. Mol. Gen. Genet. 177:399–412.
59. Kim, K. J., M. G. Sunshine, B. H. Lindqvist, and E. W. Six. 2001. Capsid size determination in the P2–P4 bacteriophage system: suppression of *sir* mutations in P2's capsid gene N by supersid mutations in P4's external scaffold gene *sid*. Virology 283:49–58.
60. Kirby, J. E., J. E. Trempy, and S. Gottesman, 1994. Excision of a P4-like cryptic prophage leads to Alp protease expression in *Escherichia coli*. J. Bacteriol. 176:2068–2081.

61. Kita, K., J. Tsuda, T. Kato, K. Okamoto, H. Yanase, and M. Tanaka. 1999. Evidence of horizontal transfer of the EcoO109I restriction-modification gene to *Escherichia coli* chromosomal DNA. J. Bacteriol. 181:6822–6827.
62. Krevolin, M. D., and R. Calendar. 1985. The replication of bacteriophage P4 DNA in vitro. Partial purification of the P4 α gene product. J. Mol. Biol. 182:509–517.
63. Krevolin, M. D., R. B. Inman, D. Roof, M. Kahn, and R. Calendar. 1985. Bacteriophage P4 DNA replication. Location of the P4 origin. J. Mol. Biol. 182:519–527.
64. Lagos, R., R. Z. Jiang, S. Kim, and R. Goldstein. 1986. Rho-dependent transcription termination of a bacterial operon is antagonized by an extrachromosomal gene product. Proc. Natl. Acad. Sci. USA 83:9561–9565.
65. Lengyel, J. A., R. N. Goldstein, M. Marsh, M. G. Sunshine, and R. Calendar. 1973. Bacteriophage P2 head morphogenesis: cleavage of the major capsid protein. Virology 53:1–23.
66. Lin, C. S. 1984. Nucleotide sequence of the essential region of bacteriophage P4. Nucleic Acids Res. 12:8667–8684.
67. Linderoth, N. A., and R. Calendar. 1991. The Psu protein of bacteriophage P4 is an antitermination factor for rho-dependent transcription termination. J. Bacteriol. 173:6722–6731.
68. Lindqvist, B. H., and E. W. Six. 1971. Replication of bacteriophage P4 DNA in a nonlysogenic host. Virology 43:1–7.
69. Lindqvist, B. H., G. Dehò, and R. Calendar. 1993. Mechanisms of genome propagation and helper exploitation by satellite phage P4. Microbiol. Rev. 57:683–702.
70. Liu, T., S. K. Renberg, and E. Haggård-Ljungquist. 1997. Derepression of prophage P2 by satellite phage P4: cloning of the P4 ε gene and identification of its product. J. Virol. 71:4502–4508.
71. Liu, T., S. K. Renberg, and E. Haggard-Ljungquist. 1998. The E protein of satellite phage P4 acts as an anti–repressor by binding to the C protein of helper phage P2. Mol. Microbiol. 30:1041–1050.
72. Marvik, O. J., E. Jacobsen, T. Dokland, and B. H. Lindqvist. 1994. Bacteriophage P2 and P4 morphogenesis: assembly precedes proteolytic processing of the capsid proteins. Virology 205:51–65.
73. Marvik, O. J., P. Sharma, T. Dokland, and B. H. Lindqvist. 1994. Bacteriophage P2 and P4 assembly: alternative scaffolding proteins regulate capsid size. Virology 200:702–714.
74. Marvik, O. J., T. Dokland, R. H. Nokling, E. Jacobsen, T. Larsen, and B. H. Lindqvist. 1995. The capsid size-determining protein Sid forms an external scaffold on phage P4 procapsids. J. Mol. Biol. 251:59–75.
75. Ow, D. W., and F. M. Ausubel. 1980. Recombinant P4 bacteriophages propagate as viable lytic phages or as autonomous plasmids in *Klebsiella pneumoniae*. Mol. Gen. Genet. 180:165–175.
76. Piazza, F., M. Zappone, M. Sana, F. Briani, and G. Dehò. 1996. Polynucleotide phosphorylase of *Escherichia coli* is required for the establishment of bacteriophage P4 immunity. J. Bacteriol. 178:5513–5521.
77. Pierson, L. S., III, and M. L. Kahn. 1984. Cloning of the integration and attachment regions of bacteriophage P4. Mol. Gen. Genet. 195:44–51.
78. Pierson, L. S., III, and M. L. Kahn. 1987. Integration of satellite bacteriophage P4 in *Escherichia coli*. DNA sequences of the phage and host regions involved in site-specific recombination. J. Mol. Biol. 196:487–496.
79. Polo, S., T. Sturniolo, G. Dehò, and D. Ghisotti. 1996. Identification of a phage-coded DNA-binding protein that regulates transcription from late promoters in bacteriophage P4. J. Mol. Biol. 257:745–755.
80. Pruss, G. J., J. C. Wang, and R. Calendar. 1975. In vitro packaging of covalently-closed circular monomers of bacteriophage DNA. J. Mol. Biol. 98:465–465.
81. Ravatn, R., S. Studer, A. J. Zehnder, and J. van der Meer. 1998. Int-B13, an unusual site-specific recombinase of the bacteriophage P4 integrase family, is responsible for chromosomal insertion of the 105–kilobase *clc* element of *Pseudomonas* sp. Strain B13. J. Bacteriol. 180:5505–5514.
82. Ravin, N. V., A. N. Svarchevsky, and G. Dehò. 1999. The anti-immunity system of phage-plasmid N15: identification of the antirepressor gene and its control by a small processed RNA. Mol. Microbiol. 34:980–994.
83. Reiter, K., H., Lam, E. Young, B. Julien, and R. Calendar. 1998. A complex control system for transcriptional activation from the *sid* promoter of bacteriophage P4. J. Bacteriol. 180:5151–5158.
84. Renberg Eriksson, S. K., T. Liu, and E. Haggård-Ljungquist. 2000. Interacting interfaces of the P4 antirepressor E and the P2 immunity repressor C. Mol. Microbiol. 36:1148–1155.
85. Sabbattini, P., F. Forti, D. Ghisotti, and G. Dehò. 1995. Control of transcription termination by an RNA factor in bacteriophage P4 immunity: identification of the target sites. J. Bacteriol. 177:1425–1434.
86. Sabbattini, P., E. Six, S. Zangrossi, F. Briani, D. Ghisotti, and G. Dehò. 1996. Immunity specificity determinants in the P4-like retronphage ΦR73. Virology 216:389–396.
87. Saha, S., E. Haggård-Ljungquist, and K. Nördstrom. 1989. Activation of prophage P4 by the P2 Cox protein and the sites of action of the Cox protein on the two phage genomes. Proc. Natl. Acad. Sci. USA 86:3973–3977.
88. Sauer, B., D. Ow, L. Ling, and R. Calendar. 1981. Mutants of satellite bacteriophage P4 that are defective in the suppression of transcriptional polarity. J. Mol. Biol. 145:29–46.
89. Shore, D., G. Dehò, J. Tsipis, and R. Goldstein. 1978. Determination of capsid size by satellite bacteriophage P4. Proc. Natl. Acad. Sci. USA 75:400–404.
90. Six, E. W. 1963. A defective phage depending on phage P2. Bacteriol. Proc., p. 138.
91. Six, E. W., and C. A. Klug, 1973. Bacteriophage P4: a satellite virus depending on a helper such as prophage P2. Virology 51:327–344.
92. Six, E. W. 1975. The helper dependence of satellite bacteriophage P4: which gene functions of bacteriophage P2 are needed by P4? Virology 67:249–263.

93. Six, E. W., and B. H. Lindqvist. 1978. Mutual derepression in the P2–P4 bacteriophage system. Virology 87:217–230.
94. Six, E. W., M. G. Sunshine, J. Williams, E. Haggård-Ljungquist, and B. H. Lindqvist. 1991. Morphopoietic switch mutations of bacteriophage P2. Virology 182:34–46.
95. Souza, L., R. Calendar, E. W. Six, and B. H. Lindqvist. 1977. A *trans*activation mutant of satellite phage P4. Virology 81:81–90.
96. Strack, B., M. Lessl, R. Calendar, and E. Lanka. 1992. A common sequence motif, -E-G-Y-A-T-A-, identified within the primase domains of plasmid-encoded I- and P-type DNA primases and the α protein of the *Escherichia coli* satellite phage P4. J. Biol. Chem. 267:13062–13072.
97. Sullivan, J. T., and C. W. Ronson. 1998. Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. Proc. Natl. Acad. Sci. USA 95:5145–5149.
98. Sun, J., M. Inouye, and S. Inouye. 1991. Association of a retroelement with a P4-like cryptic prophage (retronphage ΦR73) integrated into the selenocystyl tRNA gene of *Escherichia coli*. J. Bacteriol. 173:4171–4181.
99. Swenson, D. L., N. O. Bukanov, D. E. Berg, and R. A. Welch. 1996. Two pathogenicity islands in uropathogenic *Escherichia coli* J96: cosmid cloning and sample sequencing. Infect. Immun. 64:3736–3743.
100. Tamanoi, F., H. Saito, and C. C. Richardson. 1980. Physical mapping of primary and secondary origins of bacteriophage T7 DNA replication. Proc. Natl. Acad. Sci. USA 77:2656–2660.
101. Terzano, S., R. Christian, F. H. Espinoza, R. Calendar, G. Dehò, and D. Ghisotti. 1994. A new gene of bacteriophage P4 that controls DNA replication. J. Bacteriol. 176:6059–6065.
102. Tocchetti, A., S. Serina, S. Terzano, G. Dehò, and D. Ghisotti. 1998. Identification of two replicons in phage-plasmid P4. Virology 245:344–352.
103. Tocchetti, A., G. Galimberti, G. Dehò, and D. Ghisotti. 1999. Characterization of the *oriI* and *oriII* origins of replication in phage-plasmid P4. J. Virol. 73:7308–7316.
104. Tocchetti, A., S. Serina, I. Oliva, G. Dehò, and D. Ghisotti. 2001. Cnr interferes with dimerization of the replication protein α in phage-plasmid P4. Nucleic Acids Res. 29:536–544.
105. Wang, S., P. Palasingam, R. H. Nokling, B. H. Lindqvist, and T. Dokland. 2000. In vitro assembly of bacteriophage P4 procapsids from purified capsid and scaffolding proteins. Virology 275:133–144.
106. Woelker, B., and W. Messer. 1993. The structure of the initiation complex at the replication origin, *oriC*, of *Escherichia coli*. Nucleic Acids Res. 21:5025–5033.
107. Yeo, H.-J., G. Ziegelin, S. Korolev, R. Calendar, E. Lanka, and G. Waksman. 2002 Phage P4 origin-binding domain structure reveals a mechanism for regulation of DNA-binding activity by homo- and heterodimerization of winged helix proteins. Mol. Microbiol. 43:855–867.
108. Ziegelin, G., E. Scherzinger, R. Lurz, and E. Lanka. 1993. Phage P4 α protein is multifunctional with origin recognition, helicase and primase activities. EMBO J. 12:3703–3708.
109. Ziegelin, G., N. A. Linderoth, R. Calendar, and E. Lanka. 1995. Domain structure of phage P4 α protein deduced by mutational analysis. J. Bacteriol. 177:4333–4341.
110. Ziegelin, G., R. Calendar, D. Ghisotti, S. Terzano, and E. Lanka. 1997. Cnr protein, the negative regulator of bacteriophage P4 replication, stimulates specific DNA binding of its initiator protein α. J. Bacteriol. 179:2817–2822.
111. Ziegelin, G., R. Calendar, R. Lurz, and E. Lanka. 1997. The helicase domain of phage P4 α protein overlaps the specific DNA binding domain. J. Bacteriol. 179:4087–4095.
112. Ziermann, R., and R. Calendar. 1990. Characterization of the *cos* sites of bacteriophages P2 and P4. Gene 96:9–15.

27

Bacteriophage λ and its Genetic Neighborhood

ROGER W. HENDRIX
SHERWOOD CASJENS

Bacteriophage λ (lambda) was discovered at the dawn of molecular biology as an inadvertent byproduct of studies of the genetics of *Escherichia coli* K-12, where λ resides as a prophage (221). Lambda has never been independently isolated from nature a second time, but because of its particularly fortuitous choice of time and place to make its appearance, it became one of the very small number of experimental systems that were used in elucidating our most basic understanding of biological organisms at the molecular level. It has been said for many years, and may still be true, that more scientist-years per base pair have been devoted to understanding the biology of phage λ than is the case for any other organism, and this is due only in part to the modest size of its genome (48,503 bp).

Research on phage λ has made major contributions to studies of regulation of transcription, regulatory circuitry, mechanisms of recombination of both the homologous and site-specific varieties, mechanisms of DNA replication, genetic transduction, cell lysis, virion structure and assembly mechanisms, viral evolution, and DNA sequencing and cloning technologies, to name just a few highlights. Each of these topics has been the subject of one or more major reviews, and there have been two books devoted exclusively to λ biology (142, 156). It is therefore not our aim to reproduce here all of the accumulated wisdom from more than 50 years of phage λ research.

Instead, we will give a necessarily abbreviated overview of λ's life cycle and genetic functions. This will be followed by a more detailed consideration of a topic that has received increasing attention recently and one for which our present level of understanding is made possible only by the availability of genomic sequence data for a number of related phages. That topic is the genetic and evolutionary relationships between λ and a group of similar phages, sometimes referred to as the lambdoid phages and most recently comprehensively reviewed by Campbell (43). These phages have for the most part not been studied in the same detail as phage λ, but the data that are available for them—DNA sequence as well as biochemical and genetic data—provides a new dimension to the accumulated data about λ itself. It becomes possible in this way to see λ not simply as a laboratory phenomenon but as a member of a natural population of phages, interacting and evolving in the natural environment.

References were chosen in this chapter to allow entry into the literature and may not always credit those who made discoveries. The relative emphasis given to different topics unavoidably follows the authors' expertise and interests.

A λ Overview

Biological Overview

Virions

The heads of phage λ virions (figure 27-1) are icosahedrally symmetric, isometric, and about 60 nm in diameter. Its non contractile tails are 150 nm long, and they appear slightly flexible based on how they lie on an electron microscope sample grid. The main body of the tail (the "shaft" or "tube") consists of stacked protein disks which join to the base of the conical "tail tip." This morphology—isometric head, long non-contractile tail—is the most common among the double-stranded DNA phages that have been isolated and characterized (see chapter 2 for an overview of phage classification). Phage λ has two types of tail fiber: one short fiber extending from the center of the tail tip and four long, jointed "side tail fibers," attached at about the junction between the tail shaft and the tail tip.

The double-stranded DNA chromosome, tightly packed into the head without bound proteins, is nominally 48,503 bp long (the 48,502 bp of the sequenced laboratory version

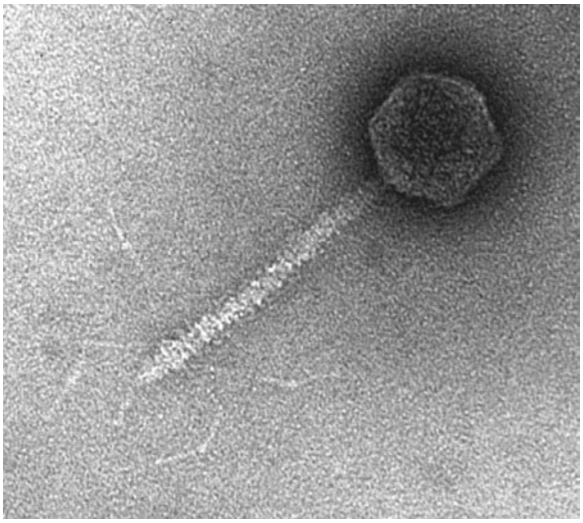


Figure 27-1 The bacteriophage λ virion. Electron micrograph of λ, negatively stained with uranyl acetate. Scale: the length of the tail, excluding fibers, is approximately 150 nm. Micrograph courtesy of Robert Duda.

of λ (314), corrected for the 1 bp deletion in the *stf* gene (see below) that was introduced in early laboratory manipulations of the phage (149)). The ends of the linear DNA molecule in the virion have 12 nucleotide single-stranded 5′-extensions; these are referred to as "cohesive ends" because of the base-pairing complementarity of the extensions at the two ends of the chromosome (155, 402).

Closer examination of virions shows that the heads have two main protein components: gpE ("gene product of gene E") and gpD, arranged in $T = 7$ *laevo* icosahedral symmetry, and present in probably exactly 405 copies each (63, 88). GpE, also known as coat protein, makes up the main structure of the shell, and the smaller gpD subunits are clustered as trimers bound to the exterior at the 3-fold and quasi-3-fold sites in the gpE lattice, where they strengthen the capsid. Quantitatively "minor" components of the head include the portal (also called the "head–tail connector"), a 12 gpB-subunit ring structure that occupies a position at only one of the 5-fold symmetric "corners" of the shell (the corner to which tails will join) (210, 353), about six copies each of gpW and gpFII, which provide a structural transition between the portal and the tail (46, 267), and about five copies each of two similar proteins ("X1" and "X2") that are constructed during capsid maturation and consist of a covalent fusion between a portion of major capsid subunit, gpE, and a portion of putative protease, gpC (147).

The virion's tail shaft consists of 32 stacked hexameric rings of the major tail subunit, gpV, corresponding to the morphologically defined disks (38, 40, 63, 195). The lumen of the tail tube is largely filled with about six copies of a proteolytically processed form of the tail-length tape measure protein, gpH (145, 354), though the right end of the DNA has descended a short distance into the head-proximal end of the tail (66, 350). The tail tip contains a few molecules each of the products of genes *J*, *L*, and *M* (and perhaps *I* and *K*) (189, 195). Of these, gpJ is present in about three copies in the mature virion and constitutes the central tail fiber and a substantial portion of the conical tip. The head-proximal end of the tail has a few copies of gpU, probably a hexameric ring on top of the stack of gpV rings that make up the shaft (197, 199). The last gene in the tail assembly pathway, *Z*, is required not for joining tails to heads but for joining them productively (56); it is not clear whether gpZ is present in mature virions.

Lytic and Lysogenic Life-styles.

As a temperate phage, λ has two alternative life-styles or "growth cycles" available to it. In the lytic cycle, the phage infects the cell by inserting its DNA into the cytoplasm. There it is circularized. An orderly expression of phage genes ensues, with the result that the phage uses the energy of the host's metabolism and its biosynthetic machinery to produce 50–100 progeny virions. The expression of certain genes eventually causes cell lysis and the release of progeny phages to infect new cells and repeat the cycle. This process is described in more detail below.

The lysogenic cycle starts, like the lytic cycle, by infection, but the majority of the phage genes become repressed soon after infection by the action of a phage-encoded repressor, the product of the *cI* gene. (Note to the λ novice: The "*I*" in the gene name "*cI*" is a Roman numeral one (182). Thus the name of the gene, as well as the name of the repressor protein, is pronounced "see-one." Pronouncing it as "see-eye" is an error, and doing so in public may elicit ribald laughter and ridicule from those who know better.) During establishment of the repressed state, the λ chromosome becomes integrated into the continuity of the bacterial chromosome. Once this occurs, the phage DNA, now known as a "prophage," is replicated passively and distributed to daughter cells as a part of the host chromosome.

A cell carrying a λ prophage is said to be a "λ lysogen" or to be "lysogenic for λ." This association can persist for an indefinite number of generations with no apparent harm to the host (95, 227), because all the lytic genes of the prophage are repressed. This life-style can be considered a "cycle" because, after an indeterminate period as a quiescent prophage, the prophage DNA can become "induced" to lytic growth mode and make progeny phage. Induction results from loss of repression by the CI repressor (see below). Discussion of lysogeny in general, and regulation of the lambda life-cycle in particular, can also be found in chapters 7–9.

Repressor and Lysogenic Conversion

The cytoplasm of a lysogenic cell has a concentration of CI repressor that is sufficient to keep the resident prophage's lytic genes repressed; as a result, if another λ virion injects its DNA into the cell, it will immediately also be repressed and therefore be unable either to enter the lytic cycle or to make the enzyme (integrase) that would allow it to become a part of the host chromosome and ensure its replication in the lysogenic cycle. A λ lysogen is thus said to be immune to infection by λ. Prophage immunity is specific to the phage since it is based on the specific binding between the repressor and its cognate operators. Thus, phage 434, a close relative of phage λ, is unaffected by the presence of a λ prophage in a cell it infects because of the different specificity of binding between its repressor and operators, and conversely λ can infect a 434 lysogen.

The change in the phenotype of a lysogenic cell that results from the expression of the repressor by the prophage—that is, immunity to infection by a phage with the same repressor specificity—is a specific example of the more general phenomenon of lysogenic conversion. Five other genes in λ are known to be expressed from an otherwise repressed prophage and as a result to change the phenotypic properties of the host cell; in other words, they cause lysogenic conversion. These include the *rexA* and *rexB* genes (263, 351), whose products block successful infection by several phages including *rII* mutants of phage T4 (and which thereby played a central role in some of the classic early experiments of molecular biology; 24, 25), the *sieB* gene (283), whose product blocks infection by a different group of phages, and two genes, *lom* and *bor*, which are thought to make the bacterial host a more effective pathogen of mammals. In the case of *lom* this increase in pathogenicity occurs by helping the lysogen bind to mammalian cells and in the case of *bor* this occurs by making the lysogen more resistant to killing by serum (17, 358).

Chromosome

Figure 27-2 is a physical map of the λ genome, showing the genes as boxes, the various DNA sites regulating transcription and other aspects of the phage's life-cycles. The arrows indicate the directions and locations of the major transcripts. The orientation of the map, with head and tail genes on the left and the early genes (i.e., those expressed early in the lytic cycle) on the right, is the conventional representation. The region from the left end to *attP*, containing the head and tail genes, is called the "left arm" and the region from *attP* to the right end, containing the early genes, is called the "right arm." The genes are for the most part organized into large operons, grouped by function and by when their expression is needed. This organization facilitates control of gene expression by regulating a small number of promoters.

The ends of this map correspond to the ends of the DNA molecule in the virion. When the phage injects its DNA into the cell upon infection, the DNA ends join to make a circle. The cohesive ends are regenerated toward the

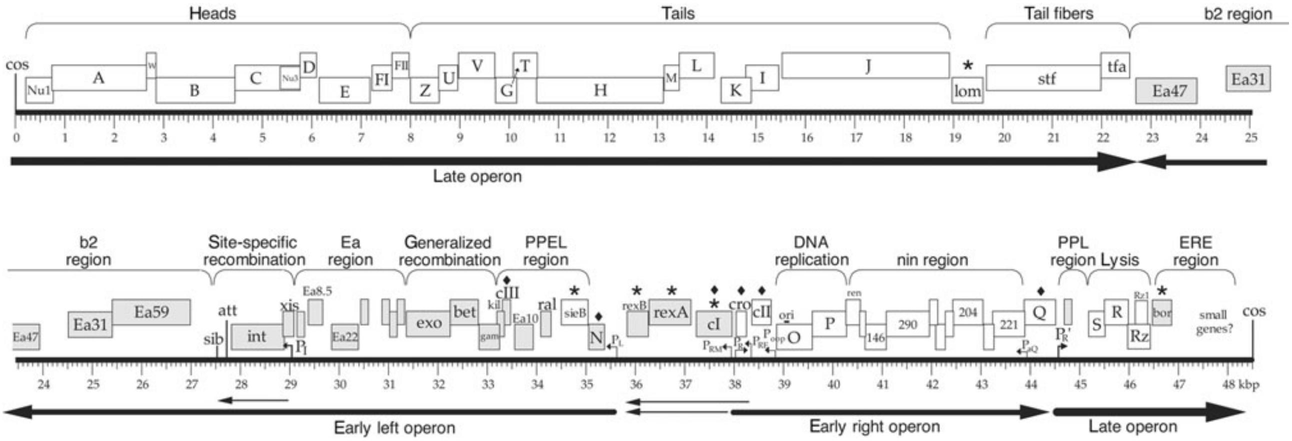

Heads
Tails
Tail fibers
b2 region
cos
Nu1
A
W
B
C
Nu3
D
E
FI
FII
Z
U
V
G
T
H
M
L
K
I
J
*
lom
stf
tfa
Ea47
Ea31
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Late operon
b2 region
Site-specific recombination
Ea region
Generalized recombination
PPEL region
DNA replication
nin region
PPL region
Lysis
ERE region
Ea47
Ea31
Ea59
att
sib
int
xis
P_I
Ea8.5
Ea22
exo
bet
gam
kil
cIII
Ea10
ral
sieB
N
P_L
rexB
rexA
cI
cro
cII
P_RM
P_R
P_RE
P_oop
ori
O
P
ren
146
290
204
221
Q
P_aQ
P_R'
S
R
Rz1
Rz
bor
small genes?
cos
24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 kbp
Early left operon
Early right operon
Late operon

Figure 27-2 The bacteriophage λ chromosome. Features of the bacteriophage λ chromosome are shown in a diagrammatic and somewhat simplified form. Genes are represented by rectangles (white, transcribed rightward; gray, transcribed leftward; vertical positions are varied for graphical clarity). Selected gene names are given on or above these rectangles. Asterisks (*) denote genes that are expressed in the lysogen and diamonds (♦) denote regulatory genes. A scale in kilobase pairs is shown below the genes, and below this the major characterized transcripts are indicated by black arrows. Important promoters are indicated by small arrows, and other important DNA sites are marked by vertical lines, just above the kilobase pair scale. The slanted arrow between the G and T reading frames indicates a programmed translational frameshift (see text). The region names b2, Ea, PPEL (promoter proximal early left), nin, PPL (promoter proximal late), and ERE (extreme right end) are discussed in the text.

end of the lytic cycle as DNA is packaged into new virions. On the other hand, when the phage enters the lysogenic cycle the circle is broken at a different site—the attachment site, or *attP*—during the integration of the DNA into the host chromosome (see figure 7-1 in chapter 7 for a cartoon of λ insertion). The result is that the prophage sequence is a circular permutation of the sequence shown in figure 27-2, with the two ends of the prophage corresponding to the sequences flanking the two sides of *attP*.

Overview of Molecular Mechanisms

Lytic Cycle

The lytic cycle can start either with prophage induction (described below) or with infection. The initial interaction between virion and host cell is carried out by the side tail fibers, which interact with an abundant component of the cell surface, probably the outer membrane porin OmpC (149). Subsequently the central (gpJ) tail fiber binds to the LamB protein, also an outer membrane protein that is normally involved in maltose transport (and whose synthesis is subject to catabolite repression; 404).

LamB binding triggers a presumed conformational change in the tail and release of the DNA from the virion into the cell. The DNA is preceded out of the virion by the tail-length tape measure protein, gpH, which occupies much of the lumen of the tail. Where in the cell the tape measure protein ends up, however, is not clear (307, 308). The details of how the DNA transits the two membranes and periplasm also is not understood, but the host PstM protein, an inner membrane protein and part of the mannose import system, is involved (108, 260, 317, 383). Immediately after arrival in the cytoplasm the cohesive ends of the λ chromosome anneal and are joined together by the host DNA ligase (120). This yields a covalently continuous double-stranded circular molecule.

All λ transcription is done by the host RNA polymerase. When the λ DNA first enters the cell, the polymerase initiates transcription from only two promoters with relevance to the lytic cycle: P_L and P_R (figure 27-2). (Promoters for the lysogenic conversion genes are presumably also recognized but without any known consequences for the lytic cycle.) The polymerase encounters a terminator and stops after transcribing a single gene in both cases: gene *N* from P_L and gene *cro* from P_R.

Nothing more would happen save for the action of gpN, which interacts with subsequent RNA polymerases initiating at P_L and P_R and renders them insensitive to termination signals. The N protein is, in other words, a transcription antiterminator (116, 129, 301, 319, 376). This is accomplished through an interaction among the RNA polymerase, gpN, four host proteins (NusA, NusB, NusE, and NusG), and a specific sequence in the nascent P_L and P_R mRNAs called the *nut* (N utilization) site. This mechanism requires that the *nut* site be part of the transcript being synthesized by the polymerase, and consequently it only works for promoters with an associated *nut* site (only P_L and P_R in λ). Because of the *cis*-acting nature of *nut* sites, some RNA polymerases are "antiterminated" (insensitive to terminators) at the same time that other RNA polymerases that started at different promoters are not. This difference is in fact the basis for an important part of the regulation of prophage integration and excision (described below).

The most important effect of gpN antitermination is that transcription initiated at P_L and P_R can be extended across the remainder of the early genes. From P_L, transcription extends past *N* through several additional genes (see figure 27-2). These genes include *cIII*, which has a role in the decision between lytic and lysogenic growth, and three genes, *exo*, *bet*, and *gam*, which promote homologous recombination during lytic growth (274, 332, 335). The products of the latter three genes have been studied extensively and have been shown to encode an exonuclease (212, 229), a strand annealing protein (190, 225), and an inhibitor of the host RecBC nuclease (238, 275), respectively.

From P_R the first gene to be transcribed after *cro* is *cII*, which plays a central role in the lytic/lysogenic decision (395). The *cII* gene is followed by genes *O* and *P*, whose products are essential for initiating phage DNA replication (118). The next 10 genes are not required for phage growth, although in some cases biochemical functions have been defined for them.

The last gene in the early transcript from P_R is gene *Q*. Gene *Q* protein is responsible for turning on high level expression of the late genes, which it does by allowing RNA polymerase at the late promoter P_R' to escape from a strong pause site (306). Like gpN, gpQ interacts with a special site near the promoter to alter the RNA polymerase, but it differs from gpN in that gpQ interacts with the site in the DNA and not with the corresponding site in the nascent RNA transcript, as is the case for gpN (303). The overall functional effect is the same in that RNA polymerase that initiates at P_R' and interacts with gpQ is insensitive to termination signals and continues to be insensitive through the entire 26.3 kbp of the late gene operon.

The "early period" of λ lytic growth ends as expression of the late genes begins, 10–12 minutes after infection. Of the many activities encoded by the early genes, the only ones that have essential roles in the lytic cycle in the laboratory are those encoded by *N*, *cro*, *O*, *P*, and *Q*. The Cro protein is a repressor which, like CI repressor, binds to the operators O_L and O_R, each of which has three subsites, and each subsite binds a single repressor dimer. The O_L and O_R operators control P_L and P_R, respectively (134, 281). However, the "rules" for how the two repressors bind to the operators (affinity, cooperativity, order of binding to subsites) are different, and the two repressors consequently have different quantitative effects on expression from P_L and P_R. The precise biological role of these operator subsite differences remains unclear (231, 300). The Cro protein builds up during the early period of lytic growth and toward the end of that period causes a decrease, though not a complete shut-off, of transcription from P_L and P_R.

Replication of phage DNA begins soon after the beginning of expression from P_R, initiated by the action of gpO and gpP. The O protein binds to the replication origin, which is a tandem series of four 19 bp near-exact repeats (called "iterons") lying within the gene *O* coding region (101, 355). The bound O protein recruits the P protein to the origin which in turn recruits the host DnaB helicase (236). This complex is then partially disassembled by the host DnaK, DnaJ, and GrpE chaperones to release DnaB from the origin (these ubiquitous chaperones were discovered and named because of their participation in λ DNA replication; 403). The remainder of the host replication apparatus is then added and replication proceeds away from the origin (87, 118, 245, 270, 408, 409). The λ O protein is metabolically unstable, so that when P_R transcription stops during establishment of lysogeny, initiation of DNA replication is also rapidly extinguished (127, 410).

During lytic growth, or prior to the establishment of lysogeny, replication initially proceeds bidirectionally from the origin to produce daughter circles from the initial circular template chromosome. At 10–15 minutes into the lytic growth cycle, replication changes to the rolling circle mechanism, which produces the multi-genome linear concatemers that are the substrate for packaging into the head. The mechanism of this switch to rolling circle replication remains unclear (15).

The end of the early period and the beginning of the late period is marked by the dramatic increase in transcription of the late gene operon from P_R', and therefore by the beginning of gpQ activity. The well-studied λ *Q* gene lies at the end of the early rightward transcript and its product, the Q protein, causes high levels of late gene transcription by facilitating release of RNA polymerase from a pause site just downstream of the late promoter, P_R'. In addition, gpQ renders the RNA polymerase insensitive to downstream *cis* termination signals by contacting its σ^{70} subunit (237, 259, 303, 306). Essentially all transcription of the late genes starts at P_R' and continues through the entire late gene operon. Because it takes RNA polymerase more than 10 minutes to traverse the operon, the time when synthesis of the various late proteins begins depends on their position in the operon (294). As far as is known, these differences in onset of translation have no functional significance. The late operon genes encode the proteins involved in building progeny virions and in cell lysis.

Virion Assembly: Overview

Assembly of the structural components of the virion depends on ordered interactions among the structural proteins and not on their order of synthesis. The amount of messenger RNA synthesized is essentially the same for all of the late operon genes, and the stability of that messengerRNA is nearly the same whether measured by a physical or functional assay (294, 295). Despite this, the molar yield of protein from the different genes varies over a nearly 1000-fold range. The translational efficiency for a given gene is determined by the region surrounding the beginning of the gene—presumably the translation initiation signals (311). The translational yields for the different genes appear to correlate not with the genes' positions in the operon but with the amount of each protein that is needed for its biological function, in most cases virion assembly (57). When the various proteins required for virion assembly and concatemeric DNA accumulate to suitable levels, assembly of progeny virions occurs via a specific assembly pathway—a set of ordered protein–protein and protein–DNA interactions (122, 195). Like other tailed phages, λ assembles heads and tails independently, and then these join to form functional virions (371, 372).

Head Assembly

Like most other large viruses, λ assembles its head by first assembling a procapsid—an empty coat protein shell—and then actively pumping the DNA chromosome into that shell (184, 185) (see chapter 6 for a general overview of the packaging of dsDNA in phage heads). λ has 10 genes that are required for head assembly, and the products of six of them end up in the mature virion (table 27-1). Only a subset of the head genes, *B, C, Nu3* and *E*, are needed for procapsid assembly. Twelve subunits of gpB make up the grommet-like portal structure. The hole in the center of this dodecameric ring is the "portal" through which DNA will enter the procapsid. The GroEL/S host chaperonin system is needed to produce functional portals, probably at the level of subunit folding (indeed, it was the study of this process in λ that led to the initial discovery of this chaperone, and gave it its name; (9, 123)).

Once portals are made, they may act as a nucleus for procapsid assembly, although this has not been rigorously shown, and some procapsid-like particles assemble in the absence of gpB (reviewed by 122, 251). The details of the procapsid assembly pathway are not clearly understood, but aspects of some roles of individual proteins are known. The gene *C* protein has a particularly complex and interesting role. It is tentatively identified as the head maturation protease because of its sequence similarity to the ClpP protease family and because of the protein composition of procapsids and other phage-related aberrant structures that are assembled when gpC is defective. Gene *Nu3* encodes the putative scaffolding protein; scaffolding proteins aid the proper assembly of coat proteins by co-assembling with them in the interior of the procapsid, but they are removed from the structure before DNA is packaged.

Table 27-1 Virion Structure and Assembly Genes of Bacteriophage λ

Gene	Size of protein (amino acids)	In virion?	Copies	Function, special features
Nu1	181	N		Small terminase subunit: DNA packaging
A	641	N		Large terminase subunit: DNA packaging
W	68	Y	~6	Adaptor between portal and gpFII
B	533	Y[a]	12	Portal
C	439	Y[b]	~10	Protease
Nu3	131	N		Scaffolding protein
D	110	Y	405	Major capsid decoration protein
E	341	Y	405	Major capsid subunit
FI	132	N		Accessory role in DNA packaging
FII	117	Y	~6	Forms tail attachment site on head
Z	192	?		Head–tail assembly
U	131	Y	~6	Tail shaft stabilization
V	246	Y	192	Major tail subunit
G	140	N		Tail assembly chaperone
T	279[c]	N		Extension, by translational frameshift, of gpG tail assembly chaperone
H	853[d]	Y	~6	Tail length tape measure protein
M	109	Y?		Tail tip assembly
L	232	Y?		Tail tip assembly
K	199	N?		Tail tip assembly
I	223	N?		Tail tip assembly
J	1132	Y	~3	Tail tip assembly, central tail fiber
stf	774	Y	12	Side tail fiber, main structural component
tfa	194	Y	12	Side tail fiber, assembly factor and structural component

[a] Twenty-one amino acids cleaved from N-terminus of most subunits.
[b] Processed into X1 and X2.
[c] Size given is for G-T frameshift product.
[d] Approximately 100 amino acids removed during tail maturation.

The *Nu3* gene is nested in-frame in the last one third of the *C* gene; this means that the C-terminal one third of the gpC sequence is identical to the gpNu3 sequence (323). The functional implications of this sequence relationship are not well understood, but because multiple scaffolding proteins typically occupy the coat shell interior by themselves, it is tempting to speculate that the C-terminal portion of gpC co-assembles with the gpNu3 molecules that make up the scaffold. (We could see a gpC–gpNu3 interaction either as a way of assuring that gpC gets included into the assembly or as a mechanism by which gpC, already assembled to the portal, could nucleate assembly of the scaffold.)

Regardless of how gpC comes to be part of the structure, once the procapsid is assembled, each of the approximately 10 copies of gpC becomes covalently bonded to one of the 415 copies of gpE that initially form the procapsid shell, and the gpC–gpE "fusion product" gets trimmed proteolytically to make two slightly different sized products named X1 and X2 (147). (Numerology: It would take 420 copies of gpE to make a complete T = 7 shell. In the actual prohead five of these are missing to make space for the portal, giving 415 gpEs. Of these, approximately 10 get converted to X1 or X2 through reaction with gpC, leaving 405 gpEs in the mature prohead and head.) The locations of X1 and X2 in the procapsid structure are not known, but a plausible model suggests that they are located around the portal where they act as structural adaptors between the regular gpE lattice (that makes up the procapsid shell) and the portal (which interrupts that regular lattice; 252). Other proteolytic events that are part of procapsid maturation include removal of 21 amino acids from the N-terminus of the portal protein, gpB (though curiously only from about two–thirds of the subunits) (144, 367), and fragmentation of the scaffolding protein, gpNu3, which is then lost from the structure (144, 168, 293).

Following assembly and proteolytic maturation, the procapsids are ready to package a phage chromosome. In addition to the procapsids, packaging requires the replicated DNA in the form of a multigenome concatemer and three head proteins that are not part of the mature virion, the products of genes *Nu1*, *A*, and *FI*. The GpNu1 and gpA polypeptides are the subunits of a multitalented enzyme known as terminase (so named because it creates the termini of the virion chromosome). Terminase recognizes and binds to the *cos* site on the concatemeric DNA, makes a staggered cut of the DNA to produce the left cohesive end of the DNA that is about to be packaged, and then carries the end of the DNA to the portal vertex of the procapsid, where the terminase docks with the portal to form the DNA packaging pump. This pump is driven by ATP hydrolysis that is probably carried out at an active site on the gpA subunit (65, 94). Following packaging of a full complement of DNA, the terminase recognizes and cuts at the downstream *cos* site and probably departs from the head bound to the DNA downstream from the cut, which is the beginning of the next chromosome that will be packaged.

The role of gpFI in this process is an auxiliary one in that phage mutants missing a functional *FI* gene still produce infectious virions, albeit with about 1% of wild-type efficiency, and *FI* independent mutants are easily isolated (254). Genetic experiments suggest that gpFI interacts with the gpA subunit of terminase and with the major capsid subunit, gpE (64, 253). A plausible model is that gpFI is part of the DNA–terminase complex and facilitates docking of that complex with the portal of a procapsid by mediating the initial contact between the complex and the procapsid.

During DNA packaging, and in some way triggered by the packaging process, the procapsid shell undergoes a dramatic rearrangement of the gpE subunits. This rearrangement produces a shell that is larger in diameter, has walls that are thinner in cross–section, is smoother surfaced, and is more angular in overall appearance (hexagonal in outline). This rearrangement exposes new sites at the points of 3-fold symmetry in the gpE lattice, and these are binding sites for the second major capsid component, gpD, which joins the capsid as trimers and significantly strengthens the structure (88, 122, 171, 340). Gene D protein's binding stabilizes the capsid substantially and in fact is necessary to make the capsid strong enough to successfully contain a full genome's worth of DNA (340). Once packaging is complete, the final two head proteins, gpW and gpFII, add to the head in that order in about six copies each (46, 58, 267). GpFII forms the binding site for the tail, which will join spontaneously.

Tail Assembly

Tail assembly in λ begins at the tip of the tail with gpJ. The products of genes *I, K, L,* and *M* then act in that order to finish the tail tip (189, 195, 198). In a parallel reaction, the products of genes *H, G,* and *G-T* form a complex which then interacts with the tail tip and the major tail subunit, gpV, to produce a tail that is complete except for the addition of two proteins at the head proximal end (193, 194). These last two proteins are gpU, which adds about six subunits to cap and stabilize the tail shaft (199), and gpZ, whose mode of action is unknown but in the absence of which tails join to heads and the resulting particles are defective in DNA injection (56).

The interaction of the products of genes *H, G,* and *G-T,* mentioned above, is a particularly interesting and central part of tail assembly. Gene *H* protein is an *α*-helical protein which is cleaved in an assembly-dependent fashion by an unidentified protease (146, 354) and which acts as a tape measure to determine the length of the tail shaft (192, 196). *G* and *T* are overlapping open reading frames that are expressed through a translational frameshift mechanism, similar to the expression of *gag* and *pol* in many retroviruses. This arrangement results in the production of a large amount of gpG and a small amount of gpG-T, the frameshift product that includes an N-terminal half with the sequence of gpG and a C-terminal half with the sequence encoded by the "*T*" reading frame (224).

GpG and gpG-T act as assembly chaperones by coating newly synthesized tape measure protein, gpH, and holding it in an extended conformation. The C-terminal "T" domain of gpG-T binds soluble gpV, the major tail shaft subunit, and most likely serves both to recruit gpV to the site of assembly and to induce it to change into its assembly-competent conformation. Thus activated, gpV is presumed to assemble around the tape measure protein to form the tail shaft, displacing and replacing the gpG and gpG-T chaperones as it does so (399).

It is evident from this description how the tape measure protein may determine tail length, namely by binding a standard (probably saturating) amount of the gpG and gpG-T assembly chaperones. Thus, if the major tail subunit gpV only assembles by interacting with and replacing the chaperones bound to the tape measure protein, the length of the tail will be determined by the length of the tape measure protein. The lengths of the tails of phages λ, HK97, and HK022, as well as those of a number of other long-tailed phages, are roughly proportional to the sizes of their tape measure proteins (264).

Cell Lysis

Progeny virions are assembled soon after late proteins appear, and cell lysis happens abruptly some 30–35 minutes after translation of the lysis genes commences. Phage λ has three genes and five proteins that are responsible for cell lysis at the end of the lytic cycle. These genes are situated at the beginning of the late gene operon, and two of them, as described below, are "double" genes in the sense that they each encode two distinct protein products.

The most straightforward of the lysis functions, the cell wall hydrolase or endolysin, is encoded by the middle gene of the group—gene *R* in λ. The λ R protein is a transglycosylase that hydrolyzes a particular bond in the cell's peptidoglycan—specifically a 1,4-β linkage between *N*-acetyl-D-glucosamine and *N*-acetyl muramic acid. This digestion weakens the cell wall sufficiently to allow the osmotic pressure difference across the cell wall and other physical stresses to burst the cell (28, 105).

Lysis timing is determined largely by interactions between the two alternative products of the *S* gene, holin and anti-holin (368, 406). Holins disrupt the inner membrane of the cell and, because the cell wall hydrolase is separated from its cell wall substrate by the inner membrane, the timing of holin action determines when the cell wall can be degraded and therefore the timing of cell lysis. The *S* gene encodes two translational starts that result in production of two membrane proteins that differ in length by only two amino acids but have opposite biological effects. The shorter protein causes efficient membrane disruption and the longer one inhibits the action of the first. It is the interaction of these two proteins as they accumulate throughout the lytic cycle that in a complex and not well understood way determines the timing of cell lysis (31, 132, 368, 406). For more detail on S-mediated λ lysis, and phage lysis in general, see chapter 10.

The third lysis gene, *Rz* in λ, encodes an accessory lysis function in that it is required for lysis or plaque formation only in the presence of high concentrations of divalent cations. It has been suggested that the Rz protein hydrolyzes a peptide bond in the peptidoglycan structure, but there is in fact no direct evidence for this claim. Whatever its function, it also requires a second protein, called Rz1, that is encoded entirely within the *Rz* gene in a different reading frame but in the same orientation (348). Mutations have been constructed in λ that selectively knock out one or the other of the *Rz* and *Rz1* open reading frames, and both mutations have the same defective phenotype, failure to lyse in the presence of high levels of divalent cations (407).

Lysogenic Cycle

The lysogenic cycle starts, like the lytic cycle, by infection of a cell by a λ virion, and the first few events of the two cycles are the same: injection and circularization of the DNA, expression of the two immediate early genes, *N* and *cro*, and action of the N protein to allow expression of the rest of the early genes. Following these initial events a "decision" is made between lytic growth, as described above, and lysogenic growth, in which the eventual outcome is a viable cell with a repressed λ prophage incorporated into the host chromosome. The two primary known determinants of which way the decision will go are the multiplicity of infection—that is, how many individual phages infect the cell simultaneously—and the physiological state of the cell. The molecular basis of the decision has been the subject of intense and prolonged study over the past 50 years (96, 241, 281); we give here a somewhat abbreviated and simplified picture.

The proximal effector of this decision is the concentration of CII protein in the cell (157, 231, 395). CII protein is a phage-encoded transcription factor that binds to TTGCN$_6$ TTGC DNA sequences in the −35 regions of three phage λ promoters and activates those promoters from sites that are not recognized by the polymerase into promoters that sponsor a high level of transcription (160, 164, 324, 345, 395). The primary effect of CII protein is mediated through the P$_{RE}$ promoter, which, in the presence of high levels of CII protein, efficiently produces a transcript of the *cI* repressor gene. The resulting CI repressor protein effectively shuts down the lytic cycle by binding to its operators, O$_L$ and O$_R$, and preventing transcription from P$_L$ and P$_R$.

The second promoter activated by CII protein is P$_I$, which produces a transcript encoding the integrase (1). Thus when the phage DNA is heading for a repressed state because of high levels of CII protein, enough integrase is made to ensure that the repressed prophage will be successfully incorporated into the host chromosome. Also, since the *xis* gene is not expressed from the P$_I$ transcript and Xis is only required for excision (see below), the proper direction of the reaction is ensured, that is, toward lysogeny.

CII protein also activates the P$_{aQ}$ promoter, which makes an antisense RNA repressor of the *Q* gene, which delays entry into the late period of the lytic cycle (163). Thus, high levels of CII protein push the phage into the lysogenic pathway. If CII protein levels are low then the phage defaults to the lytic cycle.

The CII protein can be regarded as the phage's environmental sensor through which it determines whether the lytic or lysogenic pathway will be more beneficial to the phage's interests. The precise manner in which the host physiology affects λ's lytic/lysogeny decision remains somewhat uncertain, but CII is a metabolically unstable protein, and its concentration in the cell is therefore determined by a balance between rates of synthesis and of degradation (68, 161). On the side of synthesis, the rate of production is higher when P$_R$ is highly expressed. When there are more DNA templates in the cell—that is, at higher multiplicity of infection—lysogeny is favored, perhaps because it takes more repressor to shut down more copies of P$_R$ and/or there are more *cII* gene copies in the cell to make a higher concentration of CII (e.g., 20).

Stability of the *cII* portion of the P$_R$ messenger RNA is also regulated. P$_{oop}$ transcription, which is at least partly controlled by the host LexA SOS repressor, creates an antisense RNA to the *cII* portion of the P$_R$ transcript (213). The RNA duplex formed with the *cII* mRNA is inactivated by RNaseIII cleavage (214). Stability of the P$_{oop}$ transcript, in turn, is affected by polyadenylation by the host PcnP protein (346, 390). Finally, the host protein IHF (integration host factor) has been reported to affect *cII* gene translation (234), and recently reported effects of guanosine tetraphosphate (ppGpp) and the host SeqA, ClpP/ClpX, and DnaA proteins on lysogenization frequency of λ and/or on transcription from P$_R$ indicate that multiple signals may impinge on CII synthesis and thus on λ's lysis/lysogeny decision (81, 125, 279, 330, 331).

Proteolytic degradation of CII protein, which is also important in determining its intracellular concentration, is accomplished primarily by the host HflB (also known as FstH) membrane-bound protease system (209, 325). At high cAMP concentrations the stability of CII protein is enhanced and lysogeny is favored; conversely with low cellular cAMP concentrations, as when the cells are growing on glucose, lytic growth is favored, presumably because HflB activity is higher and CII protein degradation is faster (20, 131, 180, 309, 331) (this catabolite repression effect may be stronger in some lambdoid phages than others; 169, 285). Finally, host proteins HflA and HflD affect the levels of CII protein by their modulatory effects on the activity of HflB protease, opening the possibility that other, as yet unknown signals acting through these proteins might affect the lysis/lysogeny decision (202, 203). The phage-encoded CIII protein, which is made from the P_L transcript, also has a critical role in this balance of activities; it inhibits (and is degraded by) the activity of HflB and so favors lysogeny (153, 211).

The body of knowledge about the biochemical mechanisms by which λ lytic/lysogenic decisions are made, as sketched above, is extensive and varied. Yet it is clear that the complete story has not been told. A recent example is that it has been shown that CI repressor can form an octamer that links O_R to O_L by a DNA looping mechanism. This mechanism appears to have a crucial role in turning off P_{RM} (see below) and allowing for an effective transition into lytic growth upon prophage induction (86), and it may also have a role in the lytic/lysogenic switch. This and other recent findings about how this paradigmatic genetic switch functions are still being worked into our understanding of the process, and we can expect informative new aspects of the workings of the switch to continue to be revealed for some time to come.

Integration and Its Regulation

λ's well-studied prophage integration functions map near the center of the vegetative map. Integrase (Int), assisted by host factor, IHF, catalyzes a site-specific reciprocal recombination reaction between *attP*, the phage attachment site located just downstream of the *int* gene, and *attB*, the bacterial attachment site situated between two genes in the *E. coli* chromosome (44, 79, 217, 375). The result is the insertion of the prophage DNA into the continuity of the bacterial chromosome, circularly permuted relative to the virion DNA, and flanked by two hybrid attachment sites.

Excision, which is the macroscopic if not microscopic reversal of the integration reaction, requires the phage Xis and host Fis proteins in addition to Int and IHF (12, 69, 103). The direction of the reaction is controlled during different parts of the phage life cycles by controlling the ratio of Int to Xis to drive the reaction in the direction appropriate to the biological situation. This is accomplished in part by differential stability of the two proteins (222, 374) and by activation of the P_I promoter by CII, but primarily by an elegant "retroregulation" system in which the phage senses whether it is integrated or not and regulates Int and Xis synthesis accordingly (97, 268).

The phage senses its state of integration by whether the *sib* regulatory site, which lies across the attachment site from *int* in the nonintegrated genome, is downstream from the *int* and *xis* genes or, as in the integrated prophage, is not. A key to this regulation lies in the fact that *int* can be transcribed from two promoters, P_L and P_I. Transcripts from these promoters differ in whether they include *xis* (P_L does and P_I does not), in whether the transcribing RNA polymerase has been antiterminated by the N protein (P_L has, P_I has not), and consequently in what kind of RNA structures are formed if the polymerase encounters *sib* (P_L forms a messenger RNA destabilizing structure, P_I forms a stable 3′ end). The result of these differences is that transcripts from the two promoters have very different functional stabilities, but only if the phage is not integrated (that is, only if *sib* is included in the transcript). The overall effect is that expression from an induced, integrated prophage produces both Int and Xis, therefore allowing excision. Expression from a nonintegrated chromosome during establishment of lysogeny produces only Int, favoring integration.

Repressed State and Induction

In the repressed prophage the only genes that are expressed are the *cI* gene encoding the CI repressor and the few other lysogenic conversion genes. Because the *cII* gene is one of the genes that is firmly repressed in this condition, there is no CII protein to activate transcription of the *cI* gene from the P_{RE} promoter. Instead, *cI* is transcribed from a different promoter, P_{RM}, which is located just upstream from the *cI* coding region, overlapping the O_R operator. Like P_{RE}, expression of P_{RM} requires a transcription activator, but in this case the activator is not the CII protein but the CI protein itself (134, 165). At the concentrations of repressor typically found in a lysogen, two of the three repressor-binding subsites of O_R are occupied by CI dimers, and in this configuration P_{RM} is activated. However, if the concentration of CI repressor rises sufficiently to cause occupation of the third, lower affinity subsite of O_R, then RNA polymerase is denied access to P_{RM} and further transcription stops. Transcription is thus inhibited until cell growth decreases repressor concentration to the point that the third O_R sub site again becomes free and further *cI* transcription can commence.

Induction of the prophage into the lytic cycle happens when repression by the CI repressor is lost. In a small fraction of the population (perhaps about one cell in 10^6 per cell generation under typical laboratory growth conditions), induction occurs "spontaneously"—that is, without obvious provocation. The mechanism(s) responsible for spontaneous induction is not known, but it may simply result from the stochastic loss of functional repressor molecules to below the level needed to keep P_{RM} activated.

Treatment of the culture with an appropriate dose of ultraviolet light or other DNA-damaging agents such as mitomycin C leads to simultaneous induction of essentially every cell that is lysogenic for λ in the culture. Relative to spontaneous induction, this situation is much better understood: DNA damage-mediated induction works through the cellular SOS response. Just as the activated RecA protein that is produced in response to DNA damage turns on expression of the SOS genes by causing proteolytic autocleavage of the LexA repressor, it also causes autocleavage of the CI repressor of the phage (230, 305). The cleaved repressor is no longer able to dimerize and as a result it can no longer bind effectively to the operators. The lytic cycle thus commences.

Comparative Lambdoid Phage Genomics

What Is a Lambdoid Phage?

It became clear early in the study of λ that some of the *E. coli* phages that had been isolated independently, and which were also coming under study, seemed similar to λ in their overall temperate life-style, in their genetic map, in their virion morphology, etc. These properties could be contrasted with the strikingly different properties of some other groups of phages, notably T4 and the six other "type" phages of *E. coli*, whose study was advocated most famously by Max Delbrück (42 these "type" phages are reviewed in this volume: T1 in chapter 17; T2, T4, and T6 in chapter 18; T3 and T7 in chapter 20; and T5 in chapter 19). The λ-like, or lambdoid phages included prominently a collection of phages isolated in Paris and studied initially by the Pasteur group. Genetic hybrids between these and λ were used to great effect in early studies aimed at understanding the nature of phage immunity (gene regulation and regulator specificity), host range, and other fundamental topics of phage biology (141, 154).

To the extent that the term "lambdoid" was ever defined formally, it included the idea that a lambdoid phage was capable of recombination with λ itself to produce a functional hybrid phage, as was first shown with phage 434 by Kaiser and Jacob (183). The layout of genes along the genetic map also turned out to be largely conserved within this group. Furthermore, it became evident fairly early that the lambdoid group was not confined to phages of *E. coli*, the most obvious example being the very well studied *Salmonella enterica* phage P22 (342) (see chapter 22 for a review of phage P22 biology).

The widespread perception of phage λ as an exemplar of all temperate phages has sometimes led to usages of the term "lambdoid" that go far beyond the original intentions of those who coined the term, effectively rendering it meaningless. Thus, "lambdoid" has frequently been used as a synonym for "temperate," and it has in other circumstances been used to indicate that a phage under discussion has the same virion morphology as λ, particularly the long, noncontractile tail. Such usages remove any discriminatory descriptive power from the term. At the same time, advances in comparative genomics of the tailed phages in general, and specifically of λ and its relatives (as described in more detail below), have led to a better understanding of the genetic structure of these phage populations. This has meant, paradoxically, that it has become increasingly difficult to find a biologically meaningful sharp division between the lambdoid phages and other tailed phages (219).

Despite these difficulties, we have chosen for a more detailed discussion a group of four of the best studied lambdoid phages, whose complete genome sequences are known and which span much of the diversity in what might be considered a reasonable and current definition of the lambdoid phages. Our purpose is not to belabor the question of how to define a lambdoid phage, but rather to choose a representative selection of phages in the genetic neighborhood of λ. Comparing these phages gives a much enriched understanding of the many functions that their genomes share and that are often accomplished in subtly, or sometimes in dramatically different ways by the different phages. In addition to λ, our "comparison group" phages include *E. coli* phages HK97 and N15 (the latter is reviewed in chapter 28) and *Salmonella* phage P22 (181, 265, 291, 314) (chapter 29). We provide a detailed comparative map of these four phages' genomes (figure. 27-S1), available on the associated website at www.thebacteriophages.org.

In addition to these four, there are many other λ-like phages that infect enteric bacteria that could just as well have been included in such a comparison. Phages HK022, HK620, Sf6, SfV, 933W, VT2-Sa, Φ27, Fels-1, ST64T, ϕKO2, ES18, Gifsy-1 and -2, for example, have been completely sequenced (6, 52, 61, 181, 242, 247, 269, 296, 405, M. Pedulla, S. Casjens, and R. Hendrix, unpublished data) and specialized aspects of phages 21, 82, 434, ϕ80, PA-2, PY54 and numerous others have been studied. In addition, phages D3, ϕE125, and APSE-1 that infect nonenteric proteobacteria *Pseudomonas aeruginosa, Burkholderia thailandensis* and an endosymbiont of pea aphids (215, 359, 389), respectively, as well as largely intact prophages in several proteobacterial genome sequences, have apparent transcriptional cascades and genome organizations that are very similar to λ's, though less overt sequence similarity. We will refer to phages outside of our comparison group when they provide informative examples.

Mosaic Relationships and Their Evolutionary Origin

Lambdoid Phage Genome Mosaicism

Before we discuss the differences and similarities among these phages, we will first indicate how those relationships are thought to have arisen. This we will do by sketching our best current understanding of the mechanisms by which the tailed phages evolve, and how those mechanisms have led to the genomes of contemporary phages (see chapter 4 for additional discussion of phage evolution). This topic was first investigated in the late 1960s by comparing lambdoid phage genome sequences through electron microscopic visualization of DNA heteroduplexes assembled from pairs of lambdoid phages (159, 327). The striking observation from these studies was that the ability of the two DNA strands from two different phages to form a heteroduplex—a measure of nucleotide sequence similarity—varies in a patchwork fashion across the lengths of the genomes. This argues that the genomes are genetic mosaics, generated by non-homologous (or possibly site-specific) recombination in the ancestry of the phages. These putative sites of recombination appeared to lie preferentially at certain locations, and it was hypothesized—and largely confirmed when the genomic sequences became available—that these sites are located at gene boundaries.

Such observations led to the “modular evolution” model (11, 33, 342), which states that the horizontal exchange of genetic modules (genes or groups of genes) is mediated by homologous recombination between genomes at specific sites located between modules, and that this mechanism generates genomes with novel combinations of genes and thereby novel lambdoid phages. Such a mechanism, as well as recombination between homologous genes, no doubt contributes to rapid shuffling of alternative modules among mosaically related phages, and in some phages nearly ubiquitous sequences—now called “boundary sequences”—do appear to be present in some intergenic locations (71). The modular evolution model, however, does not explain how new mosaic junctions (novel sequence joints) arise nor how new non-homologous genes enter the phage gene pool.

The analysis that is made possible by complete DNA sequences of multiple lambdoid genomes allows a much more detailed and subtle view of probable evolutionary mechanisms than was possible from the heteroduplex data. The most significant change in our views is that it appears that much and probably most of the non-homologous recombination that produces new mosaic joints is not confined initially to the positions of module boundaries. Rather, it is likely that these recombination events occur quasi-randomly, both with respect to position along the genome and with respect to alignment of the recombining genomes (181). The expected result is a mélange of recombinant types, the great majority of which are not packagable into virions or are otherwise defective as phages and thus immediately lost to natural selection. The recombinants that do survive will be strongly biased toward those that do not disrupt important functions. This means that in most cases they occurred at a module boundary.

In some cases, it appears that the recombination event giving rise to a surviving recombinant was slightly out of register, giving rise to a short quasi-duplication (143, 181). Such out-of-register events (both deletions and duplications) may be much more common than is apparent from observable sequences. For example, if a duplicated sequence does not provide a selective advantage it should be susceptible to removal by subsequent deletion, giving rise to what would then appear to be the product of an in-register recombination event.

In any event, non-homologous recombination between phages generates variation in genotypes in the population and therefore provides grist for the mill of natural selection. Such recombination events probably happen relatively infrequently, but given the astonishingly huge population sizes of tailed phages (estimated at $\sim 10^{31}$ individuals globally (26, 378, 387); see also chapter 33) and their probable ancient origins, the total number of such events submitted to the scrutiny of natural selection to date is most likely astronomical. Another source of variation, of course, is point mutation. Mutations will gradually accumulate in diverging genes to the point where detectable sequence similarity is lost. Thus, even genes that encode proteins that have no amino acid sequence similarity can share a common ancestry (i.e., can be homologous). A third source of genome variation is homologous recombination, inasmuch as this will rapidly reassort through the phage population the novel sequences created by the first two means (above). The number of different extant lambdoid genomes is not known. However, the variety is thought to be extremely large, and in fact identical lambdoid phages have not been independently isolated from nature, even when multiple lambdoid phages were isolated from a relatively restricted location (159).

Phage Morons

The mechanisms discussed immediately above create new genomes by rearranging and modifying the sequences of existing genomes, but they do not introduce novel functions to the genome. A possible means of doing so has been identified in the form of small genetic elements—usually one or a small number of genes flanked by a promoter and a transcription terminator—that have been inserted in recent evolutionary time between two genes that are adjacent in a comparison phage (150, 181). These elements, which have been termed "morons" ("units of more DNA"), in some cases express proteins of known function, and in these cases the genes are sometimes active from an otherwise repressed prophage and provide a function that appears to be beneficial to the host. The λ late operon moron genes fit the definition of lysogenic conversion genes (17, 358).

It need not be true that all morons contain lysogenic conversion genes, and even such genes that appear to benefit the host may sometimes only do so when the cell lyses (364, 365). Thus, we postulate that although moronic DNA may be added randomly to phage genomes rather frequently on an evolutionary time scale, most such additions are lost because they provide no selective benefit to or overtly damage the phage. Of the others, some are retained because they directly benefit lytic growth of the phage and others because they benefit the phage indirectly by acting to give selective advantage to the lysogenic host—again the huge population size of these phages make such intrinsically improbable events palatable. This view of morons as generalized units of addition to genomes has been extrapolated speculatively back in evolutionary time to suggest that, in principle, the entire phage genome could have been built by a stepwise addition of morons (150).

At this point it is unclear what the biochemical mechanism is that causes insertion of a moron into a genome. Any novel DNA sequence, including a moron, can in principle be inserted into a genome by "random" non-homologous recombination (DNA joining) events. Alternatively, morons could move by an unknown, more directed mechanism.

The Bigger Picture of Phage Relationships

The complex set of mosaic relationships among the lambdoid phages described in the preceding paragraphs is embedded in an even larger population of phages that may not initially appear to have any relationship to λ. This can be illustrated by comparing bacteriophage Mu to the four lambdoid phages in our comparison set. Mu has no easily detectable sequence similarity to any of these four lambdoid phages, at either the nucleotide or amino acid sequence level, and the overall organization of genes, DNA replication life-style, and transcription patterns are significantly different between Mu and the lambdoid phages. Furthermore Mu, unlike λ, has a contractile tail. However, the recently sequenced *Shigella* phage SfV (6) and *E. coli* phage ϕP27 (296) show that the λ-like and Mu-like groups of phages, which have previously been thought to be genetically distinct, are in fact part of a larger group of phages that have been exchanging genes with each other in the fairly recent evolutionary past

The genes of phages SfV and ΦP27 are arranged like those of a conventional lambdoid phage. The early genes of SfV, for example, are related in a typically mosaic fashion to those of λ, HK97, and P22, the "*b2* region" (see below) has similar genes to those in the *b2* region of P22, and the head genes are in the HK97 head-gene sequence family. The surprise is that the SfV and ΦP27 tail genes make very good matches to those of phage Mu, and they in fact have contractile tails. Thus, this tail module appears to have been exchanged between phages that are members of otherwise quite distinct groups, and this in turn suggests the possibility that all lambdoid and Mu-like phages are partaking of the same pool of genes. The boundaries of the lambdoid phages thus are increasingly difficult to define. A curious taxonomic consequence of their different types of tail module is that different members of the lambdoid phage group would be or are formally classified into each of the three(!) families of tailed phages—the *Siphoviridae* (λ), *Podoviridae* (P22) and *Myoviridae* (SfV)—since the current taxonomic scheme emphasizes tail morphology (2, 219) (see chapter 2 for a review of phage classification).

Gene exchange across even larger expanses of phage sequence space can be detected as well (151); for example the λ tail fiber assembly protein gene *tfa* (see below) appears to have been transferred to (or from) phage T4 so recently that the proteins from the two phages are still functionally interchangeable (121, 137). The nature of the sequence similarities that are detected among phages that infect phylogenetically very different hosts implies (we do not give the complete argument here) that gene exchange can and does occur across the entire population of tailed phages. An extreme hypothesis about the genetic structure of the global tailed phage population might be that it is a smoothly varying genetic continuum, and our apparent ability to subdivide it into biologically distinct groups is an artifact of sparse sampling of the population. We suspect that the truth lies between this extreme and the conventional view of distinct and different phage groups susceptible to simple classification. What the ultimate fate of the lambdoid phages will be, as a biologically meaningful taxonomic grouping of viruses, is not yet clear.

Tour of the Lambdoid Phage Genomes

This section is best read with figure 27-S1, found at www.thebacteriophages.org/chapters/0270.htm, available to the reader.

Overall Genome Organization

The lambdoid phages share a common genome organization, and it is within the context of this overall similarity of organization that the considerable differences among genomes are seen. The sizes of their genomes fall in a rather narrow range of approximately 40 to 60 kbp. This correlates with the observation that all lambdoid phages analyzed to date (λ, P22, HK97, and Gifsy-2) have isometric capsids with a triangulation number (T) of 7, a size that neatly packages genomes in this size range (49, 76, 88, 280, 384; J. Conway, personal communication). It remains an interesting question to what extent the evolution of these phages is constrained by their capsid geometry.

The clustering and functional order of genes is strongly conserved as well (figures 27-2 and 27-S1). The linear form of the λ map shown in the figures represents the DNA molecule in the virion, and it also corresponds to the experimentally determined genetic map. Starting from one end of the virion DNA (conventionally defined as the "left" end and corresponding to the end where DNA packaging commences) is the cluster of genes that specify the heads of the virions. These are arranged in a stereotyped order, transcribed toward the center of the genome (rightward on the standard map), and followed in turn by the tail genes and the tail fiber genes, also transcribed rightward. Together these structural genes take up approximately half of the genome, the left arm. The right arm of the genome is largely devoted to the early genes, including regulatory genes, genes encoding replication and recombination functions, and an assortment of "nonessential" or "accessory" genes that sometimes have a quantitative effect on phage progeny yield in the laboratory, are useful in particular situations, or have no known function. Transcription of the early genes diverges from a point near the middle of the right arm near the prophage repressor (*cI*) gene. Near the right end is the late promoter from which all the late genes are transcribed as a single operon. (Late transcription is perhaps most easily visualized by considering a circular version of the map, since the late operon extends across the ends of the virion DNA, which are joined by DNA ligase immediately following DNA injection. Such a circular map corresponds to the viral DNA following injection or prophage excision, and it is equivalent topologically to the multi-genome head-to-tail concatemers that predominate at the time of late transcription.) Near the beginning of the late operon are the lysis genes, followed, as transcription proceeds across the joined ends of the virion DNA, by the head, tail, and tail fiber genes. The sequences encoding the attachment site (*attP*) for prophage integration and the associated site-specific recombination functions are located in the center of the virion chromosome and, following integrative recombination, at the ends of the prophage.

In comparing different lambdoid phage genomes, there are many examples of genes in corresponding positions that are homologous—that is, they carry out the same function *and* appear to have common ancestry, as judged by sequence similarity. In any given comparison, some pairs of such genes are more similar in sequence than others, indicating different divergence times from their common ancestors (for example, the *L* genes of λ and Gifsy-2 are 66% identical, while those of λ and HK97 are only 27% identical, yet the λ integrase gene is 100% and 23% (only over the N-terminal half) identical to its HK97 and Gifsy-2 homologs, respectively). Other genes in corresponding positions can be analogous but not homologous—they carry out the same function but they do *not* appear to have a common ancestry (i.e., they belong to different sequence families).

There are also numerous genes that constitute insertions or substitutions in one genome relative to a different genome of the lambdoid group. One of the most dramatic differences of this sort occurs between phages λ and N15. The head and tail genes and the genome ends of these two phages are very similar (averaging ~56% sequence identity in encoded proteins), bespeaking an evolutionarily rather recent genetic exchange between these two lineages, but the right arms are dramatically different (291). Consistent with this divergence, there are significant differences in genome organization between these phages; each has a number of functions that are apparently missing in the other, and for the few N15 right arm genes that are recognizable as probable homologs of λ genes, the sequence similarity is weak. Thus, in phage N15 λ-like head and tail genes are joined to a group of early genes whose evolutionarily recent genetic partners are largely outside the conventional lambdoid canon. (This is an example of why it has become more rather than less difficult to define what a lambdoid phage is.)

Other big differences within our comparison group include the facts that phage P22 has a single gene encoding its tail apparatus rather than the 13 that λ has, and that P22 has an "extra" immunity region between its head and tail genes (reviewed by 272, 342). These differences can be regarded simply as a particularly dramatic analogous but non-homologous substitution in the first case or as an insertion into the context of a standard lambdoid genome in the second.

Head Genes

The order of individual genes that have the same function within the head gene cluster is strongly conserved, even in the face of mutational changes in the sequence that erase any detectable sequence similarity (e.g., 50, 53, 57). This is true not only among the lambdoid phages, as many features of this conservation are also present across the gamut of the tailed phages investigated to date. Whether this order is so strongly conserved for functional or for historical reasons is not yet clear.

Starting from the left chromosome end, the "standard" gene order is: small followed by large terminase subunit genes, portal protein gene, maturation protease gene, scaffolding protein gene, major capsid protein gene, and genes for head completion proteins. The studied lambdoid phages conserve this order faithfully, but there are some interesting variations in the actual gene structures. In addition to these, there is often a small number of additional genes, scattered among those mentioned above, whose presence and sequences are less conserved.

In the generation of mosaic genomes by non-homologous recombination, there is little evidence for survival of the products of recombination events within the portion of the head gene region that is responsible for assembly of the procapsid (portal, protease, scaffold, and coat genes). The likely explanation for this is that the proteins encoded by these genes interact so intimately that they cannot be successfully reassorted with parallel but divergent sets of genes that accomplish their interactions in even subtly different ways. On this basis, we would expect that once the sequences of two sets of head genes have drifted apart beyond a critical distance, their encoded proteins are no longer able to interact correctly, and so they will therefore continue to drift apart.

In accord with this hypothesis, there are *at least* six distinct families of lambdoid procapsid assembly genes that are exemplified by phages λ, P22, HK97, Gifsy-2, 933W, and ES18 (100, 181, 242, 269, 314, our unpublished results). Head assembly has been studied only in the first three of these six head types. These have no recognizable similarity in their coat protein primary sequences, and there are significant differences in the details of their assembly mechanisms (reviewed by 47, 122, 148). In our focus group of four lambdoid phages, three head gene types are represented by HK97, by P22, and by λ and N15 together. The HK97-like family of these genes is quite diverse, and the similar heads of ϕP27 and SfV may constitute a subgrouping within this family.

In spite of their lack of sequence similarity, there is evidence that the genes involved in head assembly in these different phages may in fact be homologous. All four of our comparison phages have a small terminase-subunit gene followed by a large terminase-subunit gene, but the effects of the terminase proteins are different in parallel with the sequence differences: both λ and N15 cut the DNA during packaging to make 12-base 5′ single-stranded cohesive ends (290, 393), HK97 makes 10-nucleotide 3′ single-stranded ends (181), and P22 cuts initially at a range of clustered sites (the “pac site” region) (60, 392) to produce a blunt end. Phage P22 subsequently cuts sequence nonspecifically as the headful packaging process moves along the DNA concatemer. Thus, within this group, the N15 and λ small terminase subunits are similar in sequence, and the P22 and HK97 subunits are unique.

Also consistent with homology in head assembly genes, the large terminase subunits are, with the portal proteins (below), the most strongly conserved (in amino acid sequence) structural proteins across a wide range of phages (50, our unpublished observations). We hasten to point out that even at this “high” level of conservation, the P22 terminase and portal amino acid sequences, for example, cannot be directly aligned with their functional counterparts in the other members of our comparison group; but they are members of a transitive set in which each member can be aligned with some but not all other members. This conservation likely reflects their central roles in the DNA recognition/cleavage process and in the DNA packaging pump (94, 328). We imagine that the complex interactions among the parts of this machine and with the phage DNA impose severe constraints on what sequence changes can be tolerated. There is a considerable body of information about how the λ terminase subunits interact with each other and with sites on the DNA (65, 83, 136). The similarities and differences with the sequences from phage N15 reinforce and augment this information (291).

Portal proteins form the hole through which DNA enters the procapsid (18, 210). They are part of the sensor that determines when the head is full of DNA and they may be active participants in the DNA translocase machine that drives DNA into the procapsid (62, 328). Also possibly relating to portal function, there are reports for several phages of a gene, of which *Bacillus* phage SPP1 gene *7* is the best example, located immediately downstream from the portal gene and encoding a protein that both interacts with the portal during virion assembly and is essential for efficient production of infectious virions (89). There is an apparent homolog of this gene in the lambdoid phage ES18 (our unpublished results) and in λ’s more distant relative, phage Mu (249). We can speculate that head assembly in most of the lambdoids has lost its need for this function, or that they never acquired this function. Alternatively and perhaps more likely, the function may be present in all the lambdoid phages but is most often incorporated into another protein such as the portal protein.

Some, but not all tailed-phages’ coat, scaffolding and portal proteins (in addition to the proteases themselves) are proteolytically cleaved by phage-encoded proteases in an assembly-dependent manner. These head maturation proteases in the lambdoid phages fall into at least two completely distinct sequence families which apparently act quite differently. The λ and N15 proteases are members of the ClpP family. About 10 copies of this putative λ protease, gpC, assemble into the procapsid where they participate in an unusual reaction in which each copy of gpC becomes covalently fused to a copy of gpE, the major capsid subunit, and each joined molecule loses about two thirds of its mass to proteolysis (147, 148). Based on indirect evidence, it is thought that gpC is responsible for this proteolysis, for the removal of 21 amino acids from the N-terminus of the portal protein, gpB, and for the degradation of the 100 or more copies copies of the scaffolding protein, gpNu3, found in the interior of the procapsid (144, 168, 251).

Phage N15 assembly has not been studied, but the similarity of the genes involved indicates that it must be very comparable to λ in this regard. In contrast, the phage HK97 protease is not recognizably part of any established protease family, though it is a member of a large family of homologous proteins found in other tailed phages. It is assembled into the procapsid in approximately 50 copies and is responsible for degradation of both HK97's putative equivalent of a scaffolding protein and of itself (90, 148). The only proteolytic processing known to occur in phage P22 head maturation is the essential cleavage of injection protein gp7 by the host OpdA protease (75), and there is correspondingly no protease gene at this position of its genome.

Scaffolding proteins occupy the interior of the procapsid, are required for proper procapsid assembly, and are removed before or during the DNA entry process. The phage P22 scaffolding protein is the prototype for this function, and it is encoded by a separate gene immediately upstream from the major capsid subunit gene (206). λ uses a different scheme for expressing its scaffolding protein in which the scaffolding protein is encoded in the last approximately one third of the protease gene by means of a strong internal in-frame translation start (323). This same arrangement for the scaffolding protein has also been found in the sequence-dissimilar phage Mu (249). The P22 and λ scaffolding proteins have no recognizable sequence homology, and in P22 they exit the procapsid intact (to be recycled in new procapsid assemblies; 59, 206) while in λ they are proteolytically destroyed. In phage HK97 the scaffolding function is thought to be fulfilled by a 102 amino acid sequence at the N-terminus of each capsid protein, and it is this sequence that is cut into pieces by the protease (93, 148). Perhaps there was a gene fusion in the ancestry of HK97 that joined the scaffolding protein and capsid subunit genes; alternatively the P22 genes may have been derived from an HK97-like ancestor by gene fission.

Sequences of coat protein (also called major capsid protein or major head protein) are very diverse, but the HK97 coat protein can be aligned with more than half the tailed-phage capsid sequences in the databases, including those of a large number of phages that infect Gram-positive hosts. Interestingly, the alignment suggests that about half these phages have the scaffolding function incorporated as an N-terminal segment of the capsid protein, as in HK97, and the rest appear to have a separate gene for the scaffolding protein just upstream from the capsid-protein gene, as in P22. Further experimental studies are required to confirm this speculation. The P22 and λ/N15 capsid proteins belong to sequence families that are not detectably related to HK97 or to each other. However, the x-ray structure of the HK97 coat protein shell and a high-resolution cryo-electron microscopy structure of the phage P22 coat protein shells argues that P22 coat protein has the same unusual fold as the phage HK97 coat protein (178, 382). In addition, there are numerous similarities in the head structures and assembly mechanisms for all three groups which, if taken together, argue that all three groups of capsid proteins may share common ancestry and may have retained the same protein fold and biochemical/functional properties in the face of the great divergence of their amino acid sequences.

The last two genes in the λ head region, located between the coat protein subunit gene and the start of the tail gene region, are *FI*, which has an accessory role in DNA packaging (64, 254) and *FII*, which encodes the last protein to join the assembling head and determines the specificity of tail attachment (46, 239); again, phage N15 has genes homologous to those of phage λ at this position. HK97 has two apparently unrelated genes in the corresponding positions that we imagine may serve the same functions, but this has not been tested.

This region of the P22 genome is unrelated to that of λ or HK97 and is more complex, with three genes that have essential roles in "head completion" (genes *4*, *10*, *26*), a gene encoding an uncharacterized assembly factor (gene *14*), and three genes encoding proteins with essential roles in DNA injection and which are probably themselves injected into the cell along with the DNA (genes *7*, *16*, *20*) (34, 172, 207, 342). The three P22 head completion genes may subsume the functions of λ *FI* and *FII*. For the three DNA injection genes, we speculate that these may carry out some of the functions surrounding infection that are accomplished by tail genes in a phage like λ, despite the fact that these three genes of phage P22 are generally thought of as head genes. Recent evidence shows that one of the P22 head completion proteins, gp4, has cell wall hydrolase activity (248). This may be analogous to the lysozyme found in the tail of phage T4 (188) and some other phages, though not yet located in the phage λ virion nor the virions of any of its close relatives.

Finally, the λ and N15 phages have two essential head genes which have no clearly corresponding genes in phages HK97 or P22; these are the head completion genes *W* and *D*. λ *W* is a small gene lying between the genes encoding the large subunit of terminase and the portal. The encoded protein binds to the head—presumably to the portal—following DNA packaging and provides a binding site for the FII protein, which in turn provides the binding site for the tail (46, 58). A high-resolution structure of gpW has been determined (240). Some other lambdoid phages, for example SfV and ΦP27, have a gene in the *W* position that encodes a small protein with a similarly high pI, but which has no sequence similarity to gpW (296). As suggested for the gene mentioned above that lies downstream from the portal gene in SPP1 and Mu but not λ, *W* may provide a function not enjoyed by many other phages, or alternatively the gpW function may be present but subsumed into another protein in phages such as HK97 and P22 that lack an obvious *W* homolog.

The λ *D* gene lies just upstream of *E*, and the D protein binds to the surface of the procapsid to form trimers on the surface, with the consequence that there is one gpD next to each gpE subunit. The gene D protein can be used as a "protein display system" and its atomic structure is known (246, 339, 400). Proteins of this type that stabilize the head shell are often called "decoration proteins" since they decorate the surface of the head. Some other viruses use this same strategy to stabilize their capsids, notably phage T4 and Herpesvirus (29, 158, 310), while others, such as P22, do not and instead rely solely on the considerable stabilization that comes from the conformational rearrangement (and expansion) of the capsid shell that all the tailed phages share. Phage HK97 and a minority of other tailed phages have no decoration proteins, but finish capsid maturation and stabilize the shell by forming covalent bonds among all the subunits of the shell, binding them together with a combination of covalent and topological links into a unified structure known as chainmail (91, 148, 271). No other factors are required for this crosslinking; the coat protein itself catalyzes the reaction (92). Such crosslinking was discovered in HK97, but is now known to occur in a number of other phages (109, 110, 124, 138, 244).

We note that even when two phages such as λ and N15 have a perfectly homologous set of head (or tail) genes, it does not mean that the encoded proteins can substitute for one another or that they function in precisely the same fashion. The relationship between the 10 homologous λ and phage 21 head genes has been particularly well explored by Feiss and colleagues (179, 326, 333, 398), who found that the 21 terminase function absolutely requires participation of the host IHF protein whereas λ terminase does not, and that among the 10 head gene products only one 21 protein, gpFII, can fully substitute for the parallel λ gene product in head assembly. This large fraction of failures to complement almost certainly reflects the intimate protein–protein interactions that take place during virion assembly.

Tail Genes

Among our four comparison lambdoid phages, the tails of P22 are very different from the others and the λ, HK97, and N15 tails are quite similar to each other in morphology, genetic organization, and, over much of the tail gene region, in sequence. The latter three have long, noncontractile tails. The tails of phage P22, by contrast, consist of six trimers of a single protein, the tailspike protein, gp9, and morphologically the tail is very short and more like a floret of protein than a tail (126). The tailspike protein has been studied quite extensively as a model for protein folding (23, 27), and a crystal structure is known, showing it to be primarily a large β-helix (336, 338 and references therein). The tailspike is also an enzyme, with endorhamnosidase activity, which allows it to degrade the external O-antigen polysaccharide of its *Salmonella* host, presumably as a means of gaining access to the cell surface for infection (175, 337) (though see chapter 29 for a counterargument for the utility of endorhamnosidase activity). There are at least two other lambdoid tail paradigms that do not happen to appear in our four comparison phages: phage 933W has an intermediate-length tail about half the head diameter and tail genes are not recognizably similar to any known genes (269); and phages ΦP27 and SfV have contractile tails that are similar to phage Mu tails (discussed above) (6, 296).

Phage λ has 11 genes devoted to making the tail proper (that is, not including the side tail fibers), and HK97 and N15 have the same number. For phage N15 these are all homologs of the λ genes based on shared sequence similarity. For phage HK97, the leftmost five tail genes are not obviously related to their λ counterparts by sequence, but for three of these five (genes *12–14*) functions have been established and match those of the λ genes in the corresponding positions (genes *V, G,* and *G-T*). The six rightmost HK97 tail genes have similar sequences to λ and N15 tail genes (λ genes *H, M, L, K, I,* and *J*).

The organization of the tail genes of these three phages is evidently the same, and it is likely, as with the head genes, that the functional order of tail genes is conserved in phages with long, noncontractile, λ-style tails even in the absence of detectable sequence similarity. Because clearly defined and unique functions have not yet been ascertained for a number of the tail genes, it is not yet possible to test this assertion as clearly with the tail genes as with the head genes. However, genes corresponding functionally to λ genes *V, G, G-T, H,* and *J* can be identified in a wide range of phages with this tail morphology, and their relative order is universally conserved (399). Mechanisms of tail assembly for phages with this type of tail have been studied in detail only for λ itself (192, 195, 196), but the similarity of most of the tail genes of phages such as HK97 and N15 to those of λ make it highly likely that they assemble tails similarly.

The roles of the virion proteins in the injection process is one of the more poorly understood aspects of lambdoid phage biology. Since their terminal tail structures are considerably smaller than the baseplates of the larger phages, such as T4, it has been much more difficult to obtain evidence of structural changes that likely occur during adsorption and injection. Nonetheless, it seems very likely that substantial "baseplate" rearrangements analogous to those that occur during T4 infection (80) do occur in all tailed phages, since the DNA has to be released from the virion only upon receptor binding. It is known that gpH, in λ, is released from the tail core (307, 308) and has some role in successful entry of DNA into the cytoplasm (98, 317). In phage P22 three proteins (gp7, gp16 and gp20, which assemble into the head rather than the tail) similarly participate in DNA entry into the cell (166, 167, 172).

Tail Fibers

The *J* gene of phage λ encodes the tail fiber that binds to the well-studied λ receptor of *E. coli*, the LamB protein of the outer membrane. The J protein constitutes much of the mass of the conical tail tip, probably as a trimer, and a portion of it extends from the end of the tail tip to form the short fiber that is visible in electron micrographs (195, 200). This fiber is most likely the part of the tail that interacts with the LamB receptor, and if so it is probably made of the C-terminal portion of gpJ since host-range mutants that affect the interaction with LamB map to the C-terminal region of gpJ (39, 74, 250, 369, 377).

Phages HK97 and N15 have homologs of λ's gpJ that are presumed to carry out this same function. It is curious that even though λ and HK97 both use the LamB receptor (85), the sequence similarity between their J proteins does not extend to their C-terminal parts, the portion of the protein thought to interact with the receptor. P22 is not known to have a fiber comparable to the J protein fiber, but the "initial contact" receptor-binding function of those fibers is incorporated into the trimeric tailspike protein, gp9 (173, 174). (The receptor is the cell-surface O-antigen lipopolysaccharide.) It is interesting to note that the nonhomologous receptor-binding proteins of λ and P22 are both trimeric and both utilize the N-terminal parts of the proteins to bind the virion (337).

Many phages have additional tail fibers of a different sort, of which the long, bent tail fibers of phage T4 are perhaps the best known. Its fibers are made up of homotrimers of different proteins that constitute the body of each half fiber: two proteins form the "elbow" and two are required for assembly of the fiber (reviewed in 388). One of the assembly-promoting proteins, gp38, may (in phage T2) or may not (in phage T4) be retained as a component of the assembled fiber (302). The fiber proteins form complex sequence families that are characterized by extensive *intra*genic mosaicism (135, 312).

As an offshoot of studies of the relationships among these tail fibers, it was discovered that phage λ also has tail fibers of this sort, although in λ the fiber is built from only one polypeptide—the Stf (s̲ide t̲ail f̲iber) protein (149). These fibers had not been known previously because the variant of λ that is used in most laboratories around the world, λ*PaPa*, has a frameshift mutation in the structural gene for the fibers. This mutant lacks the long "side tail fibers" that can be found on the version of λ that comes directly from the original *E. coli* K-12 lysogen; as a consequence, it makes larger plaques under laboratory conditions (the reason it was picked in the early days of λ genetics) but adsorbs to cells somewhat less efficiently (see chapter 5 for the explanation of why larger plaques are an expected consequence of less efficient phage adsorption).

The λ Stf protein has segments of close sequence similarity to the corresponding proteins of phages T4, P2, and T5. The λ Tfa (t̲ail f̲iber a̲ssembly) protein is a close relative of T4 gp38 (above) and the two are functionally interchangeable (121, 137). Phage HK97 has two genes (*28* and *29*) in the position corresponding to λ's *stf* and *tfa*, and they show weak similarity at the amino acid level to the λ proteins. The morphology of HK97's tail fibers is not well defined. N15 also has three functionally unassigned genes (genes *22, 23* and *25*) in the tail fiber region of the genome, and their only sequence similarity is that of phage N15 gp25 to phage HK97 gp28; since both N15 and HK97 have short brushy tail-tip fibers, perhaps these proteins form these fibers.

Morons and Terminators

Scattered throughout the genomes of these phages a few genes are present, usually with associated transcription signals, that we classify as morons (150, 181). These are most obvious in the head and tail regions, and so we discuss them here. Morons in their typical form include a small coding region flanked by a promoter and a stem-loop (factor independent) terminator. Morons are most obvious when they interrupt a stereotyped series of genes such as the lambdoid tail genes, where they may be inserted between two tail genes that are adjacent in a comparison phage (for example HK97 genes *15, 20, 22,* and *23*; 181). Morons are thought to be recent additions to the genome rather than ancestral genes that have been lost from the comparison phage, because (i) they do not fit into the functional clustering observed for most other lambdoid phage genes, (ii) similar morons often lie in different locations in different phages, and (iii) in a number of cases the G-C content of the moron DNA differs from that of the flanking regions, suggestive of recent horizontal transfer.

Where their functions are known, the morons in the late operons of the lambdoid phages all have roles that are not relevant to virion assembly. In regions of less highly conserved genes, such as in the early operons, it is more difficult to recognize recent additions, so genes cannot always be classified unequivocally as morons or as more fundamental parts of the phage genome; however the *sieB* (s̲uper i̲nfection e̲xclusion) lysogenic conversion gene, which lies backwards in the early left operon of λ and P22, is an excellent example of a possible moron in an early operon (283, 344).

An economical if speculative explanation for the patterns of morons seen across the phage population is that many and perhaps most of the genes in the present-day genome initially entered the genome as morons, and, when they found themselves retained because of the selective benefits they conferred, gradually mutated to become integrated into the regulatory circuitry of the phage and to have the sequence characteristics of the surrounding genes (150). An argument has been made for such an origin for the λ *D* gene (150, 249). A similar scenario might be invoked for the *immI* region of P22 (315, 318), which is located just upstream from the tailspike gene and is absent or replaced by other (apparently non-virion assembly) genes in the otherwise quite similar genomes of the P22-like phages L, APSE-1, HK620, ST64T, and Sf6 (61, 71, 139, 247, 359).

There are also numerous putative stem-loop transcription terminators scattered throughout the genomes of these phages. Some of these are associated with morons, and some, especially in the early control regions, can be understood as parts of the normal regulatory circuitry. Others fit neither of these cases in an obvious way. For example, apparent terminators immediately downstream of the coat protein gene are found in all four of our comparison phages. These terminators are typically located between genes and oriented in the direction of local transcription, arguing against their being in the genome by chance. It has been proposed that a function of the terminators may be to prevent potentially harmful transcription from the repressed prophage originating from morons or from elsewhere within or outside the prophage (181).

For phages like λ it is presumably not a problem to have terminators scattered throughout the lytic operons because of the transcription antitermination functions of the N and Q proteins that are expressed during lytic growth (116, 129, 306, 376). In accord with this, terminators are not generally found within lytic operons of phages that are not known to practice antitermination. On the other hand, a group of temperate mycobacteriophages that appear not to have such terminators within their operons have apparently found an alternative way to prevent transcription specifically from the repressed prophage: they have repressor binding sites ("stoperators") scattered across the genome that terminate transcription only when the repressor is bound (37). Such similarity of functional outcome achieved by different molecular mechanisms argues for the evolutionary importance of the function.

The *b2* Region

A deletion mutant of λ isolated in 1961 (201), λ *b2*, has lost nearly 6 kbp of DNA from the genome (hence known as the *b* or *b2* region) that starts at the attachment site (*attP*) and extends leftward toward the tail fiber genes. Except for some irregularities in prophage integration due to the alteration to the attachment site and loss of the *sib* site, λ *b2* is unimpaired for growth in the laboratory. This region encodes three well-expressed proteins made from the early transcript starting at the P_L promoter (140), one of which has been shown to have endonuclease activity (22), but it has not yet been possible to establish any beneficial function these protiens perform for the phage. Examination of other lambdoid phages shows that they frequently have similar lengths of lytically nonessential DNA in this region (that is, between the tail fiber genes and the attachment site), but it most often has an unrelated sequence. The impression given is of a region that is swapping segments of DNA in and out on an evolutionarily rapid time scale.

For phages such as λ where the functional role of the central (*b2*) region is not apparent, one view is that the genes in this region do confer a selective benefit on that phage, but that the benefit may apply in an ecological situation that has not been replicated in laboratory studies. An alternative view is that the encoded proteins may not necessarily confer any benefit but that there is a benefit to having DNA of any sequence here to act as a "stuffer," because (for example) this improves DNA packaging efficiency (340). The latter idea may apply for phages such as λ in which the phage packages the DNA between two specific sequences on the concatameric DNA; it is less obvious that it has any relevance to phages such as P22 that package by a headful mechanism (54, 177, 356).

Turning to our other three comparison phages, P22 has three genes in this region. In this case, although these genes have no apparent role in lytic growth, they do have clearly defined functions in host serotype conversion (O-antigen modification) and presumably benefit the phage indirectly by benefiting its lysogenic host (360). A number of *Shigella flexneri* phages carry similar lysogenic conversion genes that encode various surface polysaccharide modifying enzymes. Where these *S. flexneri* phages have been studied, they have included lambdoid phages in which the relevant genes lie in this position (6, 7, 72).

HK97 has a rather short *b2* region, encoding no proteins, but related to the λ sequence in such a way that it could have been derived from a λ-like *b2* region by a deletion of 3278 bp. Perhaps we have caught HK97 at a point in its evolutionary history when it is "waiting" for an infusion of new DNA. Phage HK022, which has an identical coat protein to HK97 and nearly identical DNA packaging proteins, has a genome that is about 1 kbp bigger than HK97's, so it seems likely that HK97 could accommodate additional DNA. HK97 also has a 51 bp segment in its abbreviated *b2* region that is derived from a non-coding portion of the type 1 Shiga toxin operon, suggesting that an ancestor of HK97 may have carried this toxin operon here (see below).

For phage N15, simple comparisons to the other lambdoid phages break down from this point rightward, as this is where phage N15's gene organization and sequences strongly diverge from the other phages in the group. The genes found at this position are homologs of *sopA* and *sopB* plasmid partition genes, which have a role in maintaining the linear plasmid that is the N15 prophage, and a homolog of bacterial *umuD* genes (291). A cleaved form of UmuD, UmuD′, is a subunit of the bacterial PolV error-prone DNA polymerase; the phage gene differs from the bacterial one first in that it does not encode the 25 amino acid N-terminal inhibitory peptide that gets cleaved off the bacterial UmuD protein during the SOS response, and second in that the other subunit of PolV, UmuC, is not encoded in the phage genome. Phage P1 (reviewed in chapter 24), another very different phage with a circular plasmid prophage, also has *sopA*, *sopB*, and *umuD′* genes, and the latter's product has been shown to functionally replace the host *umuD* gene (243).

Integration Functions

The elegant and complex mechanisms by which phage λ controls the expression of its integration functions are described above. The HK97 sequence in this region is nearly identical to λ's, and accordingly HK97 integrates into the same *attB* site as λ. Similarities between the two phages include *attP*, P_L, P_I, and *sib*, so although it has not been tested experimentally, it seems certain that HK97 regulates Int (integrase) and Xis (excisionase) expression essentially the same way λ does. In contrast, P22, like many phages, integrates its prophage not between bacterial genes but into a tRNA gene ($tRNA^{thr}$), thereby disrupting the tRNA gene but reconstituting it with the first 46 bp of the prophage, which are identical to the 3′ part of the tRNA gene (228) (other phages may similarly reconstitute protein-encoding genes when they integrate into them, for example phage 21's integration into the *icd* gene of *E. coli*; 45). The P22 Int sequence is only weakly related to that of λ, and their Xis sequences are not detectably similar (though, intriguingly, a 51 bp segment of the P22 *xis* sequence, presumably not functional, is found a short distance upstream from the *xis* gene of HK97; 181).

Given the apparently fine-tuned sophistication of the regulatory circuitry by which λ controls Int and Xis expression (above), it is perhaps surprising that P22 appears not to use this mechanism at all. P22 lacks a *sib* site and a P_I promoter, and unlike λ it has a rightward-oriented promoter, P_{aI}, a short distance upstream from *xis* (396). The details of regulation are not as well understood as in the λ case, but it is postulated that P_{aI} produces an antisense RNA that regulates expression of Int or Xis from the P_L transcript. In addition, the integrase gene of φ80 is inverted relative to λ (223) and the integrase genes of HK620 and Sf6 are on the other side of the *attP* site and so are likely to be regulated by other, as yet unknown mechanisms (71, 72).

The prophage form of phage N15 is a low-copy-number linear plasmid with covalently closed hairpin ends and thus not integrated into the host chromosome like the other prophages discussed here (291, 292) (for a review of phage N15's unusual biology, see chapter 28). Accordingly, N15 does not have an attachment site nor an integrase, but it does have features that can be considered analogous to them. When the linear N15 DNA enters the cell during infection, it circularizes, like λ through its cohesive ends. If the phage is entering the lysogenic cycle, the phage-encoded protelomerase enzyme (Ptl) binds to the DNA at the *telRL* site, located adjacent to the *ptl* gene, and makes staggered cuts in the DNA. To this point, this description is similar to what happens with λ integrase. However, in the next step Ptl joins the cut ends of the phage DNA not to bacterial DNA but to the ends of its own complementary strands, to form a linear prophage molecule with hairpin ends and a gene order that is a circular permutation of the gene order of virion DNA (84, 170, 287). The regulation of Ptl function is not understood, but the *ptl* gene and *telRL* site occupy locations that are perfectly parallel to λ's *int* and *attP*, respectively.

Leftward Early Operon "Ea" Region

In most lambdoid phages examined to date there is a region of a few kilobase pairs in the leftward early operon, upstream from the *int* and *xis* genes and downstream from the genes promoting homologous recombination, where it has been difficult to determine functions for the genes. Deletions have little or no negative effect on phage growth, and in some cases there are uncharacteristically large stretches of apparently unused DNA between genes (we call this the "Ea" region after the P22 and λ genes in this region). Failure to find functions for these genes may simply mean that appropriate tests have not been devised; and indeed recent experiments implicate four small genes in this area of the λ genome in modulation of the *E. coli* cell cycle (322). However, it may be the case for other genes in this region that they truly do not have a function and that their presence is being tolerated by the phage, at least in the short term, in the absence of any benefit to it. The reason for suggesting this is that in comparisons between phages, a number of the genes in this region appear to be made of fragments of other phages' genes from this region (see below). The appearance is that, like the central (*b2*) region on the other side of *attP* discussed above, this part of the genome is recombinationally more active than most other regions of the genome.

It is possible that non-homologous recombination events may be more frequent in the regions near *attP* because incorrect prophage excision events can give rise to functional phages that have either lost genes near *attP* or picked up bacterial genes near *attB* (the latter process is called "specialized transduction" and has been studied in some detail; 373). These remarks about the Ea region, of course, apply to the integrating members of the lambdoid group. On the other hand, recombination events may in fact not be more frequent in these "active" areas but rather recombination events here may be less likely than in other regions to put the resulting recombinant phage at a selective disadvantage and therefore cause the recombinant to be lost from the population.

In the case of HK97 this region is reduced to a single gene (gene *37*), but comparison with the largely similar HK022 suggests that HK97 gene *37* is a hybrid that could have been created from a fusion of the upstream and downstream halves of HK022 genes *37* and *32*, respectively, by the process of deleting parts of those two genes together with the four genes in between (143, 181). Two other examples of this type of relationship are the different lengths of homologs of λ *ea22* found in different phages, and the phage Sf6 gene *21*, which is a fusion of two phage 933W genes, L0065 and L0069 (61).

Homologous Recombination Functions

The phage-encoded genes responsible for homologous recombination have been studied extensively by genetic and biochemical means in both λ and P22 (256 and references therein). Although these groups of genes carry out the same overall function and are located at the same place in the genomes, they are so different in sequence, protein structure, and protein function that it seems certain they are not homologous. In λ, the early left operon genes *exo* (exonuclease), *bet* (strand annealing protein), and *gam* (anti-RecBCD protein) have been shown to cause homologous recombination (274, 332, 335), and their products have been studied in some detail (above).

The recombination genes of phage P22, which are not obvious homologs of the λ recombination genes (176), have also been studied in some detail, although their biochemical roles have not all been assigned. Here there are four recombination genes: *arf, erf, abc1,* and *abc2* (119, 176, 255, 273, 276). HK97, which in some nearby areas of the genome such as the regions around the *int* and *cI* genes is nearly identical to λ, has putative recombination genes that are homologs of P22 recombination genes *erf* and *abc2*. A striking feature of the comparison between these HK97 and P22 genes is that the *erf* genes are highly similar for the N-terminal approximately 150 codons of the 201 codon gene, at which point the similarity becomes suddenly and dramatically less, arguing that there has been a non-homologous recombination event at this point within the coding region of the *erf* gene (181).

Also consistent with a non-homologous recombination event, the phage Sf6 *erf* homolog is similar to the P22 gene only in its C-terminal portion (Sf6 also carries a bacterial-type single-strand binding protein gene in place of a P22-like *abc1* gene) (61). Such events within coding regions are seen considerably less frequently than those occurring between genes; the explanation in this case appears to be that Erf is a two-domain protein, and these recombination events occurred at the domain boundary, presumably making a protein that is still fully functional (181, 278). In the case of phage N15, there is no evidence for homologous recombination genes, so if they exist they must belong to novel sequence families.

There is at least one additional, different type of homologous recombination module that is found in some lambdoid phages but not in our comparison group. Phage Gifsy-2 has an exonuclease VIII gene (*recE*) homolog that lies at this position (242), and it has been shown that the *recE* exonuclease and adjacent *recT* strand-annealing protein genes (also at this position) from the *E. coli* K-12 defective lambdoid prophage Rac encode proteins that can replace host homologous recombination functions (113, 216, 385). Every lambdoid phage that has been studied in this regard carries some combination of exonuclease, strand-annealing protein, and RecBC modifier genes. Since Exo and Bet proteins have been shown to interact (282), it seems possible that these functions may not be able to be mixed and matched indiscriminately to give complete function.

Promoter Proximal Early Left Operon Accessory Genes

In addition to the early left genes discussed above, lambdoid phages often carry additional nonessential genes between the homologous recombination gene cluster and the early left promoter. In λ these are *kil, ea10,* and *ral,* which encode proteins that block host cell division, bind single-stranded DNA, and modulate type-I host restriction-modification systems, respectively (77, 233, 298).

SieB, although it lies inside the early left operons between genes *N* and *ral*, is transcribed from the opposite strand (283, 284). It contains a nested in-frame gene, *esc*, whose product inhibits SieB action; Esc thus allows phage P22 to escape exclusion by the SieB protein. The *sieB/esc* genes fit some of the definitions of a moron (above), but the fact that the *esc* gene is regulated by an antisense RNA repressor (284) that is processed from the early left operon message suggests that if it is a late arrival, it has become sufficiently integrated into the λ circuitry to take advantage of the P_L transcript in its regulation.

P22 has *ral, sieB,* and *kil* homologs in this position (its *kil* gene has a similar function but is not obviously homologous to that of λ) along with three other genes whose function is not clear. HK97 has two genes of unknown function at this location. The sequences of other lambdoid phages that have not been studied in the laboratory suggest that they often carry some combination of the above genes as well as other genes of unknown function. Of particular interest is the P22-like phage ASPE-1, a rather distantly related lambdoid phage, which carries a putative DNA polymerase gene in this position (359).

Control Functions

Comparative studies with different lambdoid phages in the control region around the *cI* gene predate the advent of genome sequences. The earliest examples are uses of phages with hybrid immunity regions to decipher the nature of gene regulation by *cI*. Such studies have continued to the present and we cite only a few examples.

Studies of λ CI and Cro repressors have included high-resolution structure determinations and investigations into the detailed nature of sequence-specific recognition of DNA by the repressors (4, 8, 19, 21, 73); parallel studies with the phage 434 repressors (3, 5, 70, 386) have provided valuable insights into which features of this process are general and which are specific to each phage. The various lambdoid phage CI proteins are now known to have many different operator binding specificities (immunities), but all lambdoid phages have the CI/Cro "switch."

One of the foundations to our understanding about how CI and Cro repressors interact with their operators and compete with each other for binding has been the fact that the three subsites of O_R are slightly different in sequence and interact differentially with the two repressors. It was therefore a surprise to learn that the O_R1 and O_R3 subsites of HK022 have identical sequences (261). It was shown subsequently that if these subsites in λ were changed to be identical, the phage still behaves apparently normally (231), and this has led to a rethinking—still under way—of our understanding of how the details of this regulation works.

In our comparison group, HK97 has the same immunity as λ (181), P22 has a different immunity (277, 316), and N15 has a still different immunity (232, 291). P22 also has a second prophage control region, called *ImmI* (a moron in the late operon), that encodes an antirepressor whose synthesis, in turn, is controlled by two adjacent repressor genes (*mnt* and *arc*) and an antisense RNA regulator (226, 315, 361, 394). The biological role of this antirepressor, which inactivates the prophage repressor (341) is not known. Could it be a device for making induction irreversible, a modulator of the lysis/lysogeny decision, or a modular induction system? If it is the latter, then we do not know its induction signal.

Phage N15 also encodes an antirepressor (288) in its equivalent of the λ Ea region, whose expression is controlled by a not yet fully understood RNA processing system that appears to be very similar to that of phage P4 (111, 112) (for review of phage P4, see chapter 26). It is also of interest to note that, unlike the integrating lambdoid phages, N15's repressor does not completely shut off the early left and early right operons (291); they are on at a (presumably) low level in the lysogen, probably because some of their genes are required for replication of the prophage plasmid.

The integrating lambdoid phages all appear to encode a CII-like protein at the same position as λ, but they vary quite widely in sequence suggesting that they have different operator specificities. The only three CII proteins that have been studied—those from λ, 21 and P22—all bind a similar TTGCN$_6$TTGC/T motif, yet cannot substitute for one another, suggesting additional not yet understood specificities (162, 397). The CII of HK97 is only about 34% identical to λ CII, and its target specificity has not been studied (181). Again, N15 has no apparent CII homolog (291).

Comparative studies with gene *N* have also been informative. Neely and Friedman (258) showed that the N protein of phage H-19B has different requirements for host factors than does the λ N protein, and these differences correlate with differences in the *nut* site sequences. More surprisingly, it was discovered that phage HK022 accomplishes antitermination of the early transcripts without an N protein, through a strictly RNA-based mechanism (13). At the location in the HK022 genome where an *N* gene would be expected there is a gene called *nun* that encodes an N-like protein that nonetheless has no role in affecting HK022 transcription. Instead, Nun is made from the repressed prophage and acts as a defense against infection of the lysogenic cell by other phages, such as λ (204, 208). In the presence of Nun, the λ N protein is displaced by Nun from its usual antitermination role. Nun interacts with the same host factors and phage *nut* site as does λ N protein, but instead of causing antitermination it promotes termination, thereby aborting the incipient λ infection.

In our comparison group, P22 has an N-like gene, *24*, that has analogous function but only minimal sequence similarity to λ N and a different RNA *nut* site binding specificity (114). It is not known whether N15 has an N-like activity. It has no candidate gene in the equivalent position in the early left operon, but the first gene in the early right operon encodes a λ Q-like antitermination protein (291). Regulation of the N15 early operons is not yet understood, and may be different from that of integrating lambdoid phages because of the need to leave the operons partially on in a lysogen (above).

DNA Replication

Lambda contributes two proteins to its lytic-cycle DNA replication: the products of genes *O* and *P*. The rest of the replication apparatus is host-encoded and recruited to the replication complex either directly or indirectly by the O and P proteins (270, 349). Phage P22 does not have a homolog of the λ P protein, but it does encode at the same position in the genome a homolog of DnaB, the protein that is recruited to the replication complex by gpP in λ (10, 379, 380). Thus, other than not requiring the host DnaB, the replication process appears to be the same as λ's.

Detectable similarity between λ gpO and its P22 homolog, gp18, extends only over the N-terminal portion of the protein, which is the part known to be responsible for binding the origin DNA in λ. Similarity does not extend over the C-terminal portion, the part that binds the gpP (381). HK97 has replication genes that are rather close homolog of those in P22 (181).

The other integrating lambdoid phages that have been characterized have one of these two arrangements: an origin binding protein gene with a nested origin sequence followed by either a λ gpP-like protein that recruits the host DnaB or by a P22 gp12-like DnaB homolog. However, not all of the known lambdoid replication strategies have been studied. For example, ΦP27, Gifsy-1, and Gifsy-2 carry a homolog of *E. coli dnaC*; in the first it lies between *O* and *P* position genes (296) and in the last two in the *P* position (242, 391); DnaC is a "helicase loader" (82, 220) that has very little sequence similarity to λ gpP or DnaB. Fels-1 encodes a large protein here that has weak primase similarity (242), and SfV has a gene with no homologs of known function in the *P* position (6). These appear to be "new," as yet unstudied, lambdoid replication initiation paradigms.

In replication, as for many of the other early functions, N15 is substantially different from the more conventional lambdoid phages. The one phage gene that has been implicated in replication, *repA*, is located to the left of the immunity region rather than to its right as in the other lambdoid phages (232, 286, 291). RepA is a large protein of 1324 amino acids. The *repA* gene is required both for replication of the linear plasmid prophage form of the chromosome and for lytic-cycle replication (289). It is sufficient to drive replication of a circular plasmid in the absence of any other phage genes, arguing that it is likely the only N15 gene with an essential role in phage replication and that the origin of replication is most likely within its coding region. Database searches suggest that RepA may be a multi-domain protein, with weak sequence matches to DNA primases near the N-terminus and to origin-binding helicases from animal viruses near the middle of the sequence (291).

Because the N15 head genes are very similar to those of λ it is presumed that a concatemer of the genome is produced for DNA packaging, and therefore that lytic replication may resemble that of λ. For prophage N15 replication, which has been investigated experimentally, it appears that the mechanism is that the linear, covalently closed hairpin-ended plasmid is replicated to produce a double-length, double-stranded, head-to-head circular dimer of the genome, which is resolved into two single-length linear plasmids by the action of protelomerase (48, 84, 287, 289). The nature of the switch between the very different lytic and prophage replication modes in N15 is completely unknown, but protelomerase must not function during lytic growth since *cos* cleavage combined with *telRL* cleavage would give rise to half-genome fragments of DNA.

The "Nin" Region (the "Poster Child" for Genome Mosaicism)

Between λ's DNA replication genes, *O* and *P*, and the late regulator gene, *Q*, lie 10 tightly packed genes which can be deleted with no apparent effect on phage growth under laboratory conditions. This region is usually called the "Nin region," after the *nin5* deletion that renders phage growth N-independent by removing transcription terminators that block gene *Q* transcription in the absence of gpN (78). Functions have been established for a few of the lambdoid Nin genes—protein phosphatase, homologous recombination proteins, Holliday junction resolvase, and escape from Rex exclusion—and DNA methylase and anti-repressor functions are predicted by homology, but in no case are their roles essential or clearly understood in the phages' life cycles (235, 347, 352, 362). These genes often occur in parallel positions in different lambdoid phages, and in these cases the sequences of the homologous genes are typically very similar and the differences are biased toward changes that are silent in the encoded amino acid sequence (181; R. Hendrix, unpublished observation). This observation argues that their functions are under positive selection.

Perhaps the most striking observation from comparison of the Nin region across phages is that while most lambdoid phages have such a group of genes, varying in number from one (phages Fels-1 and ΦP27) to a dozen genes (phages HK022 and Sf6), they are not always the same genes. Thus λ, HK97, and P22 all have 10 Nin region genes, but λ and HK97 share only five genes of the 10, λ and P22 share seven, and HK97, and P22 share five. The overall relationships are similar to what one would expect if each phage had "picked" its set of Nin genes from a menu of perhaps 30–50 genes (181). Because there is evidence (cited above) that the Nin region genes provide a selective benefit to the phage beyond simply occupying space in the genome, we postulate that these genes give the phage a particular constellation of tools for dealing with its environment, or, put another way, the particular set of genes that a phage has in this region contributes to its adaptation to a particular ecological niche.

Phage N15, in keeping with its many differences from the rest of the comparison group in the right arm of its genome, does not have any homologs of the genes found in the Nin regions of the other lambdoid phages. It is not yet clear whether any of the uncharacterized genes of the N15 right arm might show a similar relationship to corresponding genes of other N15-like phages.

The Q Antitermination Function and Late Operon Expression

The well-studied λ *Q* gene lies at the end of the early rightward transcript and its product, the Q protein, causes a high level of late gene transcription by facilitating release of RNA polymerase from a pause site just downstream of the late promoter, P_R'. In addition, gpQ renders the RNA polymerase insensitive to downstream *cis* termination signals by contacting its σ^{70} subunit (237, 259, 303, 306). All the lambdoid phages examined to date have an apparent *Q* homolog and, with the exception of phage N15 and its relative ϕKO2, they are all located at the same place in the gene order, just upstream from the late promoters they regulate.

The Q proteins of phages 21, 82, and λ have been studied and they form three sequence families and have three different target specificites (133, 401). The Q proteins of HK97 and P22 are both more than 96% identical to that of λ and so almost certainly have the same specificities (304). The Q proteins of phages 21, 82, and λ share only weak sequence homology, but all appear to function in the same manner (e.g., 237). The N15 *Q* homolog (gene *40*) overlaps the 3′ end of the *cro* gene homolog. In the other phages in our comparison set the space between the *cro* and *Q* genes is occupied by 13 or 14 other genes; whether this difference in genomic organization has biological significance is unknown.

The location of the late promoter in phage N15 is not known experimentally, but transcriptional timing data suggest it may lie in the space between genes *51* and *52* (291), which is well removed from the *Q* homolog but roughly at the same position in the overall gene order as is P_R' in phage λ and the others. This putative promoter is followed by a putative transcription terminator, so if the N15 Q protein acts similarly to the λ Q protein (also not tested experimentally), it could facilitate late transcription by allowing read-through of this terminator. This picture is clouded by the observation that the N15 *Q* gene is expressed from the prophage, as are a majority of the other early genes (291).

The Promoter Proximal Portion of the Late Operon

The region between the late operon promoter, P_R', and the lysis genes is only a few hundred base pairs in λ, HK97, and P22, and in N15 it contains a putative DNA adenine methylase gene; however, in some related phages it is substantially larger. Most notably, many lambdoid phages have been found to carry Shiga toxin genes in this location (191, 269, 296, 357). Shiga toxin is the most important toxin produced by enterohemorrhagic *E. coli* (130). In phages 933W and H-19B, for example, there are nearly 5 kbp between the late promoter and the lysis genes, which contain the two Shiga toxin genes and several other genes of unknown function (257, 269). The Shiga toxin genes can be classified as morons (above), but it is not yet clear whether they are regulated as late operon genes, lysogenic conversion genes, or both (364, 365).

In addition, several lambdoid phages carry what appear to be functional tRNA genes immediately downstream of the late promoter. Phages Sf6, 21, 933W, ΦP27 (as well as defective prophages, for example, 933N, 933O, 933R, and 933U in *E. coli* strain EDL933) all carry two or three putative tRNA genes in this location (61, 266, 269, 296). Sf6 and 21 carry asparagine and threonine tRNAs, while 933W and ΦP27 carry isoleucine and arginine tRNAs. The role of such tRNAs is not clear, since the codon usage of the late operons do not show any very convincing overrepresentation of the codons recognized by these tRNAs (61, 269).

Cell Lysis Genes

The λ genes *S* (holin), *R* (endolysin), and *Rz* (biochemical function unknown) lie between the late promoter and the head genes. Like phage λ, HK97 also has a transglycosylase gene at the middle of its lysis cassette, but P22 and N15 have a true lysozyme gene at this position (370). Lysozymes—which are members of a sequence family that includes hen egg white lysozyme and the lysis enzymes of other phages including T4—hydrolyze the same bond as do transglycosylases but by a subtly different enzymatic mechanism. Transglycosylases and lysozymes are not obviously homologous, but they are thought to be functionally interchangeable. Other non-lambdoid phages, including the well-studied phage T7 (chapter 20), have a third type of endolysin, an amidase, which hydrolyzes a different bond in the peptidoglycan to the same effect.

No lambdoid phages have been identified that use an amidase, but λ's cousin, phage Mu, provides an instructive example. The lysis gene of Mu encodes a lysozyme, but FluMu, an apparently intact prophage in the genome of *Haemophilus influenzae* that has a high degree of organizational and sequence similarity to Mu, carries a T7-like amidase gene in the corresponding position (249). The cell wall hydrolases thus provide a particularly clear example of genes that can participate in analogous but nonhomologous substitutions.

The holin genes of HK97 and P22 are very similar to λ *S* (51, 181). Phage 21 encodes a member of a distantly related class of holins that has been studied in the laboraotry (16), and in N15 at this position there are similarly nested genes that are predicted to encode a protein which, although it has little overt similarity in sequence to known holin proteins, has a similar pattern of hydophobic and charged amino acids (291). Likewise, phages HK97 and P22 each have clear *Rz* homologs immediately downstream of the endolysin gene position (51, 181).

Similar experiments to those in λ with genes *Rz* and *Rz1* (above) have not been done in the other lambdoid phages, but internal out-of-frame open reading frames starting, as in the λ case, with N-terminal signal sequences can be identified in each case. That observation lends support to the idea that *Rz1*-like embedded genes may have roles in cell lysis in those phages as well. N15 does not have an *Rz* homolog, but has a non-homologous gene in this position that also contains a potential signal sequence-containing internal out-of-frame gene (291).

It is not entirely clear whether the different lysis gene products interact with each other in specific ways; however, the fact that the phage T4 lysozyme and T7 Rz can replace their quite different P22 homologs without detriment in the laboratory suggests that they may function largely independently (51, 299). It is also not clear that all types of lambdoid lysis modules have been studied; the N15-like *Klebsiella* phage ϕKO2 causes the host cells to lyse yet it carries no homologs to any known lysis genes (52).

Extreme Right End of the Genome

At the extreme right end of the genome of lambdoid phages, past the lysis gene cluster, there is typically a few kilobase pairs of sequence that does not have a consistent organization from one phage to another. There are some genes in these regions with known or inferred functions, for example, the λ *bor* gene (17), a putative methylase gene in N15 (291), or the *rha* gene in P22 which inhibits successful infection of an IHF defective host (152). HK97 has an apparently intact gene, gene *73*, of unknown function in this region that has homologs at completely different locations in two other phages, D3 and 933W. Much of the rest of this region in most of the lambdoid phages examined contains no clear genes or, at best, short open reading frames of unknown functionality.

It is not clear at this point how to understand these regions. It may be that much of this sequence is non-functional in a genetic sense and is tolerated only because it is not expensive to maintain and it has not yet been deleted, or because it acts as a "stuffer" to make DNA packaging more efficient (see "The *b2* Region" above). It is tantalizing to note that like the "Ea" and "*b2*" regions (above), this region lies adjacent to a site where DNA is broken during normal λ growth; perhaps errors in COS cleavage contribute to the nature of this portion of the genome. Alternatively, it may be that it encodes beneficial functions that have not yet been recognized. A possible candidate for such a function would be small regulatory RNA molecules of the sort that have been found recently in bacterial genomes (128).

Lysogenic Conversion Gene Diversity

Many temperate phages carry genes that are expressed from the prophage and which affect the properties of the host bacterium (55, 334). As was mentioned above, λ carries six such genes: *cI*, *rexA*, *rexB* (these three are expressed at least in part as an operon), *sieB*, *lom*, and *bor*, which lie at four locations in its genome. CI, RexAB, and SieB are all able to block infection of the lysogen by other phages. Lom and Bor proteins, encoded by morons in the late operon, appear to make the lysogen more able to parasitize a mammalian host, but this has not been studied in great detail (17, 358).

In addition to repressor immunity functions, our comparison group phages each carry a different set of lysogenic conversion genes. HK97 has three late operon morons, genes *15*, *20*, and *22–23*, that may be expressed from the prophage, and two genes, *48* and *49*, that replace λ's *rexAB*, genes (181). None of these have homology to genes of known function. N15 carries a late operon moron gene, *cor*, downstream of its tail tip fiber gene, that blocks phage ϕ80 infection, and two putative DNA methylase genes on either side of the lysis gene cluster (291, 363). P22 carries, in addition to its two immunity regions, one of which is a clear late operon moron (above), two other types of phage superinfection exclusion systems, *sieB* (early left operon moron) (283) and *sieA* (late operon moron) (343), and three genes, *gtrABC*, in the "*b2*" region which encode enzymes for addition of glucose residues to the surface O-antigen polysaccharide and which are responsible for changing the antigenic properties of the host bacteria (360).

Essentially all lambdoid phages that have been examined carry lysogenic conversion genes. Such genes, if they are novel, are not always easily recognizable, but morons (which are usually lysogenic conversion genes) are unambiguously recognizable when they are inserted into a well-studied, stereotypical region such as the λ late operon. In the various sequenced late operons that have explicit homology to λ, such insertions are present in the following locations (we use the λ gene names for positional reference and list some of the phages with morons at each position): Late promoter proximal region—tRNAs (phages 21, Sf6, 933W; 61, 133, 269), Shiga toxin (933W, H-19B; 257, 269), putative DNA methylases (N15; 291), porin (PA-2; 30); within the lysis gene cluster between *R* and *Rz*—antirepressor (933W; 269); between the lysis and head genes—DNA methylase, *rha*, *bor* (N15, P22, λ; 51, 363); between *Z* and *U*—novel gene (Gifsy-1; 242); between *G-T* and *H*—novel gene (HK97; 181); between *M* and *L*—superoxide dismutase, *lom* (Gifsy-2, Fels-1; 107, 242); between *K* and *I*—novel genes (HK022, HK97, and ϕKO2; 52, 181); between *I* and *J*—superoxide dismutase, novel genes (CP-933V, HK022, HK097, Fels-1; 181, 242, 266); between *J* and *stf*—*lom*, *cor* (λ, P-EibA, 933W, N15, HK022; 181, 269, 291, 313).

No obvious moron or lysogenic conversion genes have been found inserted within the head gene cluster.

It is not known whether this is a sampling artifact or whether this region for some reason cannot tolerate such insertions, but clearly they can be tolerated in a number of different locations. It is interesting to note that *lom* and superoxide dismutase genes are found at more than one location in different phages, suggesting that they have either entered the lambdoid phages more than once or moved since their original entry. Thus, in the overall gene arrangement of the lambdoid phages, lysogenic conversion genes have been known for some time in the late operon, *b2* region, promoter proximal early left operon, and transcriptionally downstream of the prophage repressor gene. There is no obvious reason why other locations will not be found to harbor such genes, and in fact two *Salmonella* lambdoid phages, Gifsy-1 and Gifsy-2, have recently been found to harbor identical lysogenic conversion "inserts" inside the Nin region (242). This insertion contains a promoter that drives expression of two genes, a putative antirepressor and a homolog of bacterial *dinI* genes (41).

Regulation of lysogenic conversion genes can be complex. We described *sieB* and antirepressor regulation briefly above, and another case in point is the *rexAB* genes of λ, whose regulation (and mode of action) remain to be fully understood (99, 218, 329). The fact that lysogenic conversion genes affect the host but are expressed from the prophage also raises interesting possibilities for interactions with host bacterial regulatory networks. This has not been studied in many instances (e.g., regulation of the expression of λ's *lom* and *bor* genes remains unstudied); however, it is known that some are regulated by the host's SOS regulon repressor LexA (the Gifsy-1 and -2 Nin region insertion (41), and the *umuD* and *dinI* homologous genes of ϕKO2 and probably N15 (52)) and that the Shiga toxin genes on H-91B are regulated by an iron-dependent host Fur regulator protein (364). Finally, the genes encoding the phage P22 O-antigen modifying *gtrABC* genes are subject to an as yet poorly understood phase variation mechanism (117). The study of the regulation of lysogenic conversion genes should be fertile ground for future research.

Lambdoid Prophages in Nature

Any discussion of lambdoid (or any temperate) phage evolution and diversity cannot ignore the ubiquitous and plentiful prophages in bacterial genomes in nature. Surveys of the presence of prophages in Proteobacteria isolates have usually shown that a majority are able to release tailed-phage-like particles, many of which are infectious and many of which appear to be λ-like (262, 320, 321). Indeed the genome of the laboratory strain LT2 of *Salmonella enterica* serovar Typhimurium contains four functional prophages, three of which are clearly lambdoid (242). The eight currently published complete bacterial genome sequences from enteric γ-Proteobacteria genera, including specis of *Escherichia, Salmonella* and *Shigella*, contain a large number of intact and defective prophages, at least 50 of which appear to be lambdoid (50). In addition, the sequences of eight nonenteric proteobacterial genomes show that they harbor at least an additional dozen λ-like entities.

It has been reported that the environment contains three to ten tailed-phage particles for each bacterium (26, 378). Although this is not known to be the case specifically for the lambdoid phages, it seems clear that a very significant fraction of the genes of this phage group on Earth actually reside in prophages. Clearly any analysis of lambdoid phage diversity, both from the standpoint of the data available and the actual situation in nature, should include these sequenced prophages. However, this plethora of prophage data has only rather recently become available and is yet to be fully analyzed. Therefore, we will not attempt to review it in detail here, but will only mention a few aspects of the importance of prophages in lambdoid phage evolution and diversity.

It is important to realize that a majority of the "prophages" that are found in bacterial genome sequences are not complete, functional phage genomes. Many are defective prophages in various stages of mutational decay, having obviously suffered point mutations, insertions, and deletions. Some of their genes must therefore be nonfunctional. But this does not mean that such prophages do not harbor functional genes or parts of genes. For example, the *E. coli* K-12 defective lambdoid prophages, Rac and QIN, have been shown to carry functional homologous recombination genes and lysis genes, respectively (102, 187).

In Rac, it is interesting to note that the early right and late operons are largely deleted, and the genes that remain there are in a very advanced state of decay (32). Nevertheless, the early left operon (and its control) appears to be largely intact and contains, at least, functional repressor, ral, and integrase genes as well as two homologous recombination genes (36, 104, 106, 187, 205). This disparity of intactness between the two parts of the prophage could be the result of recent "repair" of the early left operon either by recombination with another prophage in the same cell or by a "passing" infecting phage. Such repair could occur through simple recombinational replacement or by integration of a second prophage to form tandem prophages followed by deletion.

We mention this to point out that recombinational generation of diversity among the lambdoid phages can occur by exchange of genetic material between a prophage and infecting phage (in either direction) or between two prophages, as well as between two phages that happen to infect the same cell. The former has been demonstrated in the laboratory for lambdoid phages in a number of situations (e.g., 102, 186), and comparison of the two quite closely related O157 *E. coli* strains EDL933 and Sakai shows that a number of inter-prophage recombinational events have almost certainly taken place since their divergence (50, see also 297). Some prophages may have very long times in residence (e.g., it may have taken millions of years for the early right and late operons of the Rac prophage to reach their current, largely decrepit state), and nearly all enteric bacteria appear to carry lambdoid prophages, so the former two scenarios seem quite likely to contribute to the recombinational generation of diversity in temperate phages, perhaps even more likely than two different but related virions simultaneously infecting the same cell.

These many experimentally unstudied prophage sequences also serve to point out additional varieties of lambdoid phage functional modules. For example, Fels-1 and Gifsy-2 (both largely unstudied functional phages, but sequenced as prophages in the *S. enterica* LT2 genome; 242) have genes in their head locations ("analogs" of λ *C* and *E*) that are unrelated or only extremely distantly related to those of the experimentally studied lambdoid phages. The Gifsy-2/Fels-1 maturation protease-coat gene, by virtue of weak domain similarities to other *non-lambdoid* phages, appears to represent a new head assembly paradigm in which the head maturation protease and coat protein are translated as one large polypeptide rather than as separate proteins. In addition, as mentioned above, Fels-1 and Gifsy-1 have replication genes that are different from those of the well-studied lambdoid phages.

This sort of observation need not be limited to prophages that are known to be fully functional. An analysis of 254 identifiable prophages in 84 completely sequenced bacterial genomes (at the time of this writing) found only a handful of convincing examples of possibly non-phage genes having been moved into prophages during the prophage decay process (50). It is thus at least provisionally justified to consider prophage genes that do not have homologs in known phages as potential "new" phage-borne genes. Perhaps the best historical example of such a situation, also mentioned above, is the *recE* homologous recombination gene of *E. coli* K-12, which resides in the defective Rac prophage. This gene was unique for many years until homologs were discovered in the "homologous recombination module" locale of the genomes of phages Gifsy-1 and Gifsy-2. In addition, it has no known non-phage homologs, so the *recE* gene family, originally discovered as part of a defective prophage, is now clearly a "phage gene."

Finally, we mention here the now well-known fact that many of the pathogenicity functions of bacteria that cause human or veterinary diseases are encoded by genes carried on prophages, both lambdoid and otherwise (14, 35, 55, 67, 366) (and see Shiga toxin discussion above as well as the broader discussions of chapter 47). We will not attempt to enumerate these here but only point out that bacteria occupy many ecological niches besides those of human pathogeneic relevance. In these cases as well it seems inconceivable that prophages are not providing genetic functions to their hosts that better adapt those hosts to their ecological situation.

Summary

Clearly, even though phage λ has been under study for over 50 years, there still remains much to be learned from studying it even further (see, for example, 115). The recent sequencing of a number of lambdoid phages that infect enteric bacteria and lambdoid prophages that reside in enteric bacterial genomes has made the field biologically much more rich and varied. The exact extent and nature of this variety remains to be fully described, and the discovery of bacterial "virulence" genes in lambdoid phage genomes has made the study of these phages more urgent due to their relevance to human health. One important question that remains is how the extensive diversity in lambdoid phages correlates with the hosts that particular phages infect. Are particular gene types more useful in some hosts than in others? Does sequence relatedness among phages with the lambdoid life-style and gene arrangement gradually dissipate as their hosts become more distantly related, or will we find abrupt changes in relatedness of phages that infect particular host groups? These and other questions regarding their evolution and diversity will only be answered as more sequence is determined and more genetic and biochemical studies are performed on additional members of the lambdoid phage group.

Acknowledgments

We thank all of our phage colleagues for the innumerable wonderful, open discussions of bacteriophage science that have always been the norm in this field. R.H.'s research is supported by NIH grants GM47795 and GM51975; S.C.'s research is supported by NSF grant MCB990526 and NIH grant AI49003.

References

1. Abraham, J., D. Mascarenhas, R. Fischer, M. Benedik, A. Campbell, and H. Echols. 1980. DNA sequence of regulatory region for integration gene of bacteriophage lambda. Proc. Natl. Acad. Sci. USA 77:2477–2481.
2. Ackermann, H. W. 1998. Tailed bacteriophages: the order *Caudovirales*. Adv. Virus Res. 51:135–201.

3. Aggarwal, A. K., D. W. Rodgers, M. Drottar, M. Ptashne, and S. C. Harrison. 1988. Recognition of a DNA operator by the repressor of phage 434: a view at high resolution. Science 242:899–907.
4. Albright, R. A., and B. W. Matthews. 1998. Crystal structure of lambda-Cro bound to a consensus operator at 3.0 Å resolution. J. Mol. Biol. 280:137–151.
5. Albright, R. A., and B. W. Matthews. 1998. How Cro and lambda-repressor distinguish between operators: the structural basis underlying a genetic switch. Proc. Natl. Acad. Sci. USA 95:3431–3436.
6. Allison, G. E., D. Angeles, N. Tran-Dinh, and N. K. Verma. 2002. Complete genomic sequence of SfV, a serotype-converting temperate bacteriophage of Shigella flexneri. J. Bacteriol. 184:1974–1987.
7. Allison, G. E., and N. K. Verma. 2000. Serotype-converting bacteriophages and O-antigen modification in Shigella flexneri. Trends Microbiol. 8:17–23.
8. Anderson, W. F., D. H. Ohlendorf, Y. Takeda, and B. W. Matthews. 1981. Structure of the cro repressor from bacteriophage lambda and its interaction with DNA. Nature 290:754–758.
9. Ang, D., F. Keppel, G. Klein, A. Richardson, and C. Georgopoulos. 2000. Genetic analysis of bacteriophage-encoded cochaperonins. Annu. Rev. Genet. 34:439–456.
10. Backhaus, H., and J. B. Petri. 1984. Sequence analysis of a region from the early right operon in phage P22 including the replication genes *18* and *12*. Gene 32:289–303.
11. Baker, J., R. Limberger, S. J. Schneider, and A. Campbell. 1991. Recombination and modular exchange in the genesis of new lambdoid phages. New Biol. 3:297–308.
12. Ball, C. A., and R. C. Johnson. 1991. Efficient excision of phage lambda from the *Escherichia coli* chromosome requires the Fis protein. J. Bacteriol. 173:4027–4031.
13. Banik-Maiti, S., R. A. King, and R. A. Weisberg. 1997. The antiterminator RNA of phage HK022. J. Mol. Biol. 272:677–687.
14. Banks, D. J., S. B. Beres, and J. M. Musser. 2002. The fundamental contribution of phages to GAS evolution, genome diversification and strain emergence. Trends Microbiol. 10:515–521.
15. Baranska, S., M. Gabig, A. Wegrzyn, G. Konopa, A. Herman-Antosiewicz, P. Hernandez, J. B. Schvartzman, D. R. Helinski, and G. Wegrzyn. 2001. Regulation of the switch from early to late bacteriophage lambda DNA replication. Microbiology 147:535–547.
16. Barenboim, M., C. Y. Chang, F. dib Hajj, and R. Young. 1999. Characterization of the dual start motif of a class II holin gene. Mol. Microbiol. 32:715–727.
17. Barondess, J. J., and J. Beckwith. 1995. *bor* gene of phage lambda, involved in serum resistance, encodes a widely conserved outer membrane lipoprotein. J. Bacteriol. 177:1247–1253.
18. Bazinet, C., J. Benbasat, J. King, J. M. Carazo, and J. L. Carrascosa. 1988. Purification and organization of the gene *1* portal protein required for phage P22 DNA packaging. Biochemistry 27:1849–1856.
19. Beamer, L. J., and C. O. Pabo. 1992. Refined 1.8 Å crystal structure of the lambda repressor-operator complex. J. Mol. Biol. 227:177–196.
20. Belfort, M., and D. Wulff. 1974. The roles of the lambda *c3* gene and the *Escherichia coli* catabolite gene activation system in the establishment of lysogeny by bacteriophage lambda. Proc. Natl. Acad. Sci. USA 71:779–782.
21. Bell, C. E., and M. Lewis. 2001. Crystal structure of the lambda repressor C-terminal domain octamer. J. Mol. Biol. 314:1127–1136.
22. Benchimol, S., H. Lucko, and A. Becker. 1982. A novel endonuclease specified by bacteriophage lambda. Purification and properties of the enzyme. J. Biol. Chem. 257:5211–5219.
23. Benton, C. B., J. King, and P. L. Clark. 2002. Characterization of the protrimer intermediate in the folding pathway of the interdigitated beta-helix tailspike protein. Biochemistry 41:5093–5103.
24. Benzer, S. 1957. The elementary unit of heredity, pp. 70–93. *In* W. McElroy, and B. Glass (eds.) The Chemical Basis of Heredity. The Johns Hopkins Press, Baltimore, MD.
25. Benzer, S. 1955. Fine structure of a genetic region in bacteriophage. Proc. Natl. Acad. Sci. USA 41:344–354.
26. Bergh, O., K. Y. Borsheim, G. Bratbak, and M. Heldal. 1989. High abundance of viruses found in aquatic environments. Nature 340:467–468.
27. Betts, S., and J. King. 1999. There's a right way and a wrong way: in vivo and in vitro folding, misfolding and subunit assembly of the P22 tailspike. Structure Fold. Des. 7:R131–R139.
28. Bienkowska-Szewczyk, K., B. Lipinska, and A. Taylor. 1981. The *R* gene product of bacteriophage lambda is the murein transglycosylase. Mol. Gen. Genet. 184:111–114.
29. Black, L. W., M. Showe, and A. C. Steven. 1994. Morphogenesis of the T4 head, pp. 218–258. *In* J. Karam (ed.) Molecular Biology of Bacteriophage T4. ASM Press, Washington, D.C.
30. Blasband, A. J., W. R. Marcotte, Jr., and C. A. Schnaitman. 1986. Structure of the lc and nmpC outer membrane porin protein genes of lambdoid bacteriophage. J. Biol. Chem. 261:12723–12732.
31. Blasi, U., and R. Young. 1996. Two beginnings for a single purpose: the dual-start holins in the regulation of phage lysis. Mol. Microbiol. 21:675–682.
32. Blattner, F. R., G. Plunkett, 3rd, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao. 1997. The complete genome sequence of *Escherichia coli* K-12. Science 277:1453–1474.
33. Botstein, D. 1980. A theory of modular evolution in bacteriophages. Ann. N.Y. Acad. Sci. 354:484–491.
34. Botstein, D., C. H. Waddell, and J. King. 1973. Mechanism of head assembly and DNA encapsulation in *Salmonella* phage P22. I. Genes, proteins, structures and DNA maturation. J. Mol. Biol. 80:669–695.
35. Boyd, E. F., and H. Brussow. 2002. Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved. Trends Microbiol. 10:521–529.
36. Brikun, I., K. Suziedelis, and D. E. Berg. 1994. DNA sequence divergence among derivatives of *Escherichia coli* K-12 detected by arbitrary primer PCR (random

amplified polymorphic DNA) fingerprinting. J. Bacteriol. 176:1673–1682.

37. Brown, K. L., G. J. Sarkis, C. Wadsworth, and G. F. Hatfull. 1997. Transcriptional silencing by the mycobacteriophage L5 repressor. EMBO J. 16:5914–5921.
38. Buchwald, M., H. Murialdo, and L. Siminovitch. 1970. The morphogenesis of bacteriophage lambda. II. Identification of the principal structural proteins. Virology 42:390–400.
39. Buchwald, M., and L. Siminovitch. 1969. Production of serum-blocking material by mutants of the left arm of the λ chromosome. Virology 42:1–10.
40. Buchwald, M., P. Steed-Glaister, and L. Siminovitch. 1970. The morphogenesis of bacteriophage lambda. I. Purification and characterization of lambda heads and lambda tails. Virology 42:375–389.
41. Bunny, K., J. Liu, and J. Roth. 2002. Phenotypes of *lexA* mutations in *Salmonella enterica*: evidence for a lethal *lexA* null phenotype due to the Fels-2 prophage. J. Bacteriol. 184:6235–6249.
42. Cairns, J., G. Stent, and J. E. Watson. 1966. Phage and the Origins of Molecuar Biology. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
43. Campbell, A. 1994. Comparative molecular biology of lambdoid phages. Annu. Rev. Microbiol. 48:193–222.
44. Campbell, A., A. del-Campillo-Campbell, and M. L. Ginsberg. 2002. Specificity in DNA recognition by phage integrases. Gene 300:13–18.
45. Campbell, A., S. J. Schneider, and B. Song. 1992. Lambdoid phages as elements of bacterial genomes (integrase/phage 21/*Escherichia coli* K-12/*icd* gene). Genetica 86:259–267.
46. Casjens, S. 1974. Bacteriophage lambda *FII* gene protein: role in head assembly. J. Mol. Biol. 90:1–20.
47. Casjens, S. 1989. Bacteriophage P22 DNA packaging, pp. 241–261. *In* K. Adolph (ed.) Chromosomes: Eukaryotic, Prokaryotic and Viral, vol. III. CRC Press, Boca Raton, Fla.
48. Casjens, S. 1999. Evolution of the linear DNA replicons of the *Borrelia* spirochetes. Curr. Opin. Microbiol. 2:529–534.
49. Casjens, S. 1979. Molecular organization of the bacteriophage P22 coat protein shell. J. Mol. Biol. 131:1–14.
50. Casjens, S. 2003. Prophages in bacterial genomics: What have we learned so far? Mol. Microbiol. 49:277–300.
51. Casjens, S., K. Eppler, R. Parr, and A. R. Poteete. 1989. Nucleotide sequence of the bacteriophage P22 gene *19* to *3* region: identification of a new gene required for lysis. Virology 171:588–598.
52. Casjens, S., E. B. Gilcrease, W. M. Huang, K. L. Bunny, M. L. Pedulla, M. E. Ford, J. M. Houtz, G. F. Hatfull, and R. W. Hendrix. 2004. The pKO2 linear plasmid prophage of *Klebsiella oxytoca*. J. Bacteriol. 186:1818–1832.
53. Casjens, S., G. Hatfull, and R. Hendrix. 1992. Evolution of dsDNA tailed-bacteriophage genomes. Semin. Virol. 3:383–397.
54. Casjens, S., and M. Hayden. 1988. Analysis in vivo of the bacteriophage P22 headful nuclease. J. Mol. Biol. 199:467–474.
55. Casjens, S., and R. Hendrix. 2005. Bacteriophages in the bacterial chromosome. *In* P. Higgins (ed.) The Bacterial Chromosome. ASM Press, Washington, D.C., pp. 39–52.
56. Casjens, S., and R. Hendrix. 1974. Comments on the arrangement of the morphogenetic genes of bacteriophage lambda. J. Mol. Biol. 90:20–25.
57. Casjens, S., and R. Hendrix. 1988. Control mechanisms in dsDNA bacteriophage assembly, pp. 15–91. *In* R. Calendar (ed.) The Bacteriophages, vol. 1. Plenum Press, New York.
58. Casjens, S., T. Hohn, and A. D. Kaiser. 1972. Head assembly steps controlled by genes *F* and *W* in bacteriophage λ. J. Mol. Biol. 64:551–563.
59. Casjens, S., and J. King. 1974. P22 morphogenesis. I. Catalytic scaffolding protein in capsid assembly. J. Supramol. Struct. 2:202–224.
60. Casjens, S., L. Sampson, S. Randall, K. Eppler, H. Wu, J. B. Petri, and H. Schmieger. 1992. Molecular genetic analysis of bacteriophage P22 gene *3* product, a protein involved in the initiation of headful DNA packaging. J. Mol. Biol. 227:1086–1099.
61. Casjens, S., D. Winn, R. Moreno, W. Inwood, and A. J. Clark. 2004. The chromosome of *Shigella flexneri* bacteriophage Sf6: complete nucleotide sequence, genetic mosaicism, and DNA packaging. J. Mol. Biol. 339:379–394.
62. Casjens, S., E. Wyckoff, M. Hayden, L. Sampson, K. Eppler, S. Randall, E. Moreno, and P. Serwer. 1992. Bacteriophage P22 portal protein is part of the gauge that regulates packing density of intravirion DNA. J. Mol. Biol. 224:1055–1074.
63. Casjens, S. R., and R. W. Hendrix. 1974. Locations and amounts of major structural proteins in bacteriophage lambda. J. Mol. Biol. 88:535–545.
64. Catalano, C., and M. Tomka. 1995. Role of gpFI protein in DNA packaging by bacteriophage lambda. Biochemistry 34:10036–10042.
65. Catalano, C. E. 2000. The terminase enzyme from bacteriophage lambda: a DNA-packaging machine. Cell. Mol. Life Sci. 57:128–148.
66. Chattoraj, D. K., and R. B. Inman. 1974. Location of DNA ends in P2, 186, P4 and lambda bacteriophage heads. J. Mol. Biol. 87:11–22.
67. Cheetham, B. F., and M. E. Katz. 1995. A role for bacteriophages in the evolution and transfer of bacterial virulence determinants. Mol. Microbiol. 18:201–208.
68. Cheng, H. H., P. J. Muhlrad, M. A. Hoyt, and H. Echols. 1988. Cleavage of the cII protein of phage lambda by purified HflA protease: control of the switch between lysis and lysogeny. Proc. Natl. Acad. Sci. USA 85:7882–7886.
69. Cho, E. H., R. I. Gumport, and J. F. Gardner. 2002. Interactions between integrase and excisionase in the phage lambda excisive nucleoprotein complex. J. Bacteriol. 184:5200–5203.
70. Ciubotaru, M., F. V. Bright, C. M. Ingersoll, and G. B. Koudelka. 1999. DNA–induced conformational changes in bacteriophage 434 repressor. J. Mol. Biol. 294:859-873.
71. Clark, A. J., W. Inwood, T. Cloutier, and T. S. Dillon. 2001. Nucleotide sequence of coliphage HK620 and the evolution of lambdoid phages. J. Mol. Biol. 311:657–679.
72. Clark, C. A., J. Beltrame, and P. A. Manning. 1991. The *oac* gene encoding a lipopolysaccharide O-antigen acetylase maps adjacent to the integrase-encoding gene on

the genome of *Shigella flexneri* bacteriophage Sf6. Gene 107:43–52.

73. Clarke, N. D., L. J. Beamer, H. R. Goldberg, C. Berkower, and C. O. Pabo. 1991. The DNA binding arm of lambda repressor: critical contacts from a flexible region. Science 254:267–270.
74. Clement, J. M., E. Lepouce, C. Marchal, and M. Hofnung. 1983. Genetic study of a membrane protein: DNA sequence alterations due to 17 *lamB* point mutations affecting adsorption of phage lambda. EMBO J. 2:77–80.
75. Conlin, C. A., E. R. Vimr, and C. G. Miller. 1992. Oligopeptidase A is required for normal phage P22 development. J. Bacteriol. 174:5869–5880.
76. Conway, J. F., R. L. Duda, N. Cheng, R. W. Hendrix, and A. C. Steven. 1995. Proteolytic and conformational control of virus capsid maturation: the bacteriophage HK97 system. J. Mol. Biol. 253:86–99.
77. Court, D., and A. B. Oppenheim. 1983. Phage lambda's accessory genes, pp. 251–277. *In* R. Hendrix, J. Roberts, F. W. Stahl, and R. Weisberg (eds.) Lambda II. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
78. Court, D., and K. Sato. 1969. Studies of novel transducing variants of lambda: dispensability of genes *N* and *Q*. Virology 39:348–352.
79. Crisona, N. J., R. L. Weinberg, B. J. Peter, D. W. Sumners, and N. R. Cozzarelli. 1999. The topological mechanism of phage lambda integrase. J. Mol. Biol. 289:747–775.
80. Crowther, R. A., E. V. Lenk, Y. Kikuchi, and J. King. 1977. Molecular reorganization in the hexagon to star transition of the baseplate of bacteriophage T4. J. Mol. Biol. 116:489–523.
81. Czyz, A., R. Zielke, and G. Wegrzyn. 2001. Rapid degradation of bacteriophage lambda O protein by ClpP/ClpX protease influences the lysis-versus–lysogenization decision of the phage under certain growth conditions of the host cells. Arch. Virol. 146:1487–1498.
82. Davey, M. J., L. Fang, P. McInerney, R. E. Georgescu, and M. O'Donnell. 2002. The DnaC helicase loader is a dual ATP/ADP switch protein. *EMBO J.* 21:3148–3159.
83. de Beer, T., J. Fang, M. Ortega, Q. Yang, L. Maes, C. Duffy, N. Berton, J. Sippy, M. Overduin, M. Feiss, and C. E. Catalano. 2002. Insights into specific DNA recognition during the assembly of a viral genome packaging machine. Mol. Cell 9:981–991.
84. Deneke, J., G. Ziegelin, R. Lurz, and E. Lanka. 2000. The protelomerase of temperate *Escherichia coli* phage N15 has cleaving-joining activity. Proc. Natl. Acad. Sci. USA 97:7721–7726.
85. Dhillon, E. K., T. S. Dhillon, A. N. Lai, and S. Linn. 1980. Host range, immunity and antigenic properties of lambdoid coliphage HK97. J. Gen. Virol. 50:217–220.
86. Dodd, I. B., A. J. Perkins, D. Tsemitsidis, and J. B. Egan. 2001. Octamerization of lambda CI repressor is needed for effective repression of P_{RM} and efficient switching from lysogeny. Genes Dev. 15:3013–3022.
87. Dodson, M., R. McMacken, and H. Echols. 1989. Specialized nucleoprotein structures at the origin of replication of bacteriophage lambda. Protein association and disassociation reactions responsible for localized initiation of replication. J. Biol. Chem. 264:10719–10725.
88. Dokland, T., and H. Murialdo. 1993. Structural transitions during maturation of bacteriophage lambda capsids. J. Mol. Biol. 233:682–694.
89. Droge, A., M. A. Santos, A. C. Stiege, J. C. Alonso, R. Lurz, T. A. Trautner, and P. Tavares. 2000. Shape and DNA packaging activity of bacteriophage SPP1 procapsid: protein components and interactions during assembly. J. Mol. Biol. 296:117–132.
90. Duda, R., K. Martincic, Z. Xie, and R. Hendrix. 1995. Bacteriophage HK97 head assembly. FEMS Microbiol. Rev. 17:41–46.
91. Duda, R. L. 1998. Protein chainmail: catenated protein in viral capsids. Cell 94:55–60.
92. Duda, R. L., J. Hempel, H. Michel, J. Shabanowitz, D. Hunt, and R. W. Hendrix. 1995. Structural transitions during bacteriophage HK97 head assembly. J. Mol. Biol. 247:618–635.
93. Duda, R. L., K. Martincic, and R. W. Hendrix. 1995. Genetic basis of bacteriophage HK97 prohead assembly. J. Mol. Biol. 247:636–647.
94. Duffy, C., and M. Feiss. 2002. The large subunit of bacteriophage lambda's terminase plays a role in DNA translocation and packaging termination. J. Mol. Biol. 316:547–561.
95. Dykhuizen, D., J. H. Campbell, and B. G. Rolfe. 1978. The influences of a lambda prophage on the growth rate of *Escherichia coli*. Microbios 23:99–113.
96. Echols, H. 1986. Bactgeriophage λ development: temporal switches and the choices of lysis and lysogeny. Trends Microbiol. 2:26–30.
97. Echols, H., and G. Gruarneros. 1983. Control of integration and excision, pp. 75–92. *In* R. Hendrix, J. Roberts, F. W. Stahl, and R. Weisberg (eds.) Lambda II. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
98. Elliott, J., and W. Arber. 1978. *E. coli* K-12 *pel* mutants, which block phage lambda DNA injection, coincide with *ptsM*, which determines a component of a sugar transport system. Mol. Gen. Genet. 161:1–8.
99. Engelberg-Kulka, H., M. Reches, S. Narasimhan, R. Schoulaker-Schwarz, Y. Klemes, E. Aizenman, and G. Glaser. 1998. *rexB* of bacteriophage lambda is an anti-cell death gene. Proc. Natl. Acad. Sci. USA 95:15481–15486.
100. Eppler, K., E. Wyckoff, J. Goates, R. Parr, and S. Casjens. 1991. Nucleotide sequence of the bacteriophage P22 genes required for DNA packaging. Virology 183:519–538.
101. Erdile, L., and R. B. Inman. 1984. The role of gene *O* protein in the replication of bacteriophage lambda. Virology 139:97–108.
102. Espion, D., K. Kaiser, and C. Dambly-Chaudiere. 1983. A third defective lambdoid prophage of *Escherichia coli* K12 defined by the lambda derivative, lambdaqin111. J. Mol. Biol. 170:611–633.
103. Esposito, D., and G. F. Gerard. 2003. The *Escherichia coli* Fis protein stimulates bacteriophage lambda integrative recombination in vitro. J. Bacteriol. 185:3076–3080.
104. Evans, R., N. R. Seeley, and P. L. Kuempel. 1979. Loss of rac locus DNA in merozygotes of *Escherichia coli* K12. Mol. Gen. Genet. 175:245–250.

105. Evrard, C., J. Fastrez, and J. P. Declercq. 1998. Crystal structure of the lysozyme from bacteriophage lambda and its relationship with V- and C-type lysozymes. J. Mol. Biol. 276:151–164.

106. Feinstein, S. I., and K. B. Low. 1982. Zygotic induction of the rac locus can cause cell death in *E. coli*. Mol. Gen. Genet. 187:231–235.

107. Figueroa-Bossi, N., and L. Bossi. 1999. Inducible prophages contribute to *Salmonella* virulence in mice. Mol. Microbiol. 33:167–176.

108. Filali Maltouf, A. K., and B. Labedan. 1985. The energetics of the injection process of bacteriophage lambda DNA and the role of the *ptsM/pel*-encoded protein. Biochem. Biophys. Res. Commun. 130:1093–1101.

109. Ford, M. E., G. J. Sarkis, A. E. Belanger, R. W. Hendrix, and G. F. Hatfull. 1998. Genome structure of mycobacteriophage D29: implications for phage evolution. J. Mol. Biol. 279:143–164.

110. Ford, M. E., C. Stenstrom, R. W. Hendrix, and G. F. Hatfull. 1998. Mycobacteriophage TM4: genome structure and gene expression. Tuber. Lung Dis. 79:63–73.

111. Forti, F., S. Polo, K. Lane, E. Six, G. Sironi, G. Dehò, and D. Ghisotti. 1999. Translation of two nested genes in bacteriophage P4 controls immunity specific transcription termination. J. Bacteriol. 181:5225–5233.

112. Forti, F., P. Sabbattini, G. Sironi, S. Zangrossi, G. Dehò, and D. Ghisotti. 1995. Immunity determinant of phage-plasmid P4 is a short processed RNA. J. Mol. Biol. 249:869–878.

113. Fouts, K. E., T. Wasie-Gilbert, D. K. Willis, A. J. Clark, and S. D. Barbour. 1983. Genetic analysis of transposon-induced mutations of the Rac prophage in *Escherichia coli* K-12 which affect expression and function of *recE*. J. Bacteriol. 156:718–726.

114. Franklin, N. C. 1985. "N" transcription antitermination proteins of bacteriophages lambda, 21 and P22. J. Mol. Biol. 181:85–91.

115. Friedman, D. I., and D. L. Court. 2001. Bacteriophage lambda: alive and well and still doing its thing. Curr. Opin. Microbiol. 4:201–207.

116. Friedman, D. I., and D. L. Court. 1995. Transcription antitermination: the lambda paradigm updated. Mol. Microbiol. 18:191–200.

117. Fukazawa, Y., and P. Hartman. 1964. A P22 bacteriophage mutant defective in antigenic conversion. Virology 23:279–283.

118. Furth, M., and S. Wickner. 1983. Lambda DNA replication, pp. 145–173. *In* R. W. Hendrix, J. Roberts, F. Stahl, and R. Weisberg (eds.) Lambda II. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

119. Garzon, A., D. A. Cano, and J. Casadesus. 1998. The P22 Erf protein and host RecA provide alternative functions for transductional segregation of plasmid-borne duplications. Mol. Gen. Genet. 259:39–45.

120. Gellert, M. 1967. Formation of covalent circles of lambda DNA by *E. coli* extracts. Proc. Natl. Acad. Sci. USA 57:148–155.

121. George, D. G., L. S. Yeh, and W. C. Barker. 1983. Unexpected relationships between bacteriophage lambda hypothetical proteins and bacteriophage T4 tail-fiber proteins. Biochem. Biophys. Res. Commun. 115:1061–1068.

122. Georgopoulos, C., K. Tilly, and S. Casjens. 1983. Lambdoid phage head assembly, pp. 279–304. *In* R. W. Hendrix, J. Roberts, F. Stahl, and R. Weisberg (eds.) Lambda II. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

123. Georgopoulos, C. P., R. W. Hendrix, S. R. Casjens, and A. D. Kaiser. 1973. Host participation in bacteriophage lambda head assembly. J. Mol. Biol. 76:45-60.

124. Gilakjan, Z. A., and A. M. Kropinski. 1999. Cloning and analysis of the capsid morphogenesis genes of *Pseudomonas aeruginosa* bacteriophage D3: another example of protein chain mail? J. Bacteriol. 181: 7221–7227.

125. Glinkowska, M., J. Majka, W. Messer, and G. Wegrzyn. 2003. The mechanism of regulation of bacteriophage lambda pR promoter activity by *Escherichia coli* DnaA protein. J. Biol. Chem. 278:22250–22256.

126. Goldenberg, D., and J. King. 1982. Trimeric intermediate in the in vivo folding and subunit assembly of the tail spike endorhamnosidase of bacteriophage P22. Proc. Natl. Acad. Sci. USA 79:3403–3407.

127. Gonciarz-Swiatek, M., A. Wawrzynow, S. J. Um, B. A. Learn, R. McMacken, W. L. Kelley, C. Georgopoulos, O. Sliekers, and M. Zylicz. 1999. Recognition, targeting, and hydrolysis of the lambda O replication protein by the ClpP/ClpX protease. J. Biol. Chem. 274:13999–14005.

128. Gottesman, S., G. Storz, C. Rosenow, N. Majdalani, F. Repoila, and K. M. Wassarman. 2001. Small RNA regulators of translation: mechanisms of action and approaches for identifying new small RNAs. Cold Spring Harb. Symp. Quant. Biol. 66:353–362.

129. Greenblatt, J., T. F. Mah, P. Legault, J. Mogridge, J. Li, and L. E. Kay. 1998. Structure and mechanism in transcriptional antitermination by the bacteriophage lambda N protein. Cold Spring Harb. Symp. Quant. Biol. 63:327–336.

130. Griffin, P. 1995. *Escherichia coli* O157:H7 and other enterohemorrhagic *Escherichia co*li, pp. 739–762. *In* M. Blaser, P. R. Smith, J. Ravdin, H. Greenberg, and R. Guerrant (eds.) Infections of the Gastrointestinal Tract. Raven Press, New York.

131. Grodzicker, T., R. R. Arditti, and H. Eisen. 1972. Establishment of repression by lambdoid phage in catabolite activator protein and adenylate cyclase mutants of *Escherichia coli*. Proc. Natl. Acad. Sci. USA 69:366–370.

132. Grundling, A., D. L. Smith, U. Blasi, and R. Y. Young. 2000. Dimerization between the holin and holin inhibitor of phage lambda. J. Bacteriol. 182:6075–6081.

133. Guo, H. C., M. Kainz, and J. W. Roberts. 1991. Characterization of the late-gene regulatory region of phage 21. J. Bacteriol. 173:1554–1560.

134. Gussin, G. N., A. D. Johnson, C. O. Pabo, and B. Sauer. 1983. Repressor and Cro protein: structure, function, and role in lysogenization, pp. 93–121. *In* R. W. Hendrix, J. Roberts, F. Stahl, and R. Weisberg (eds.) Lambda II. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

135. Haggard-Ljungquist, E., C. Halling, and R. Calendar. 1992. DNA sequences of the tail fiber genes of bacteriophage P2: evidence for horizontal transfer of tail fiber genes among unrelated bacteriophages. J. Bacteriol. 174:1462–1477.
136. Hang, Q., L. Woods, M. Feiss, and C. E. Catalano. 1999. Cloning, expression, and biochemical characterization of hexahistidine-tagged terminase proteins. J. Biol. Chem. 274:15305–15314.
137. Hashemolhosseini, S., Y. D. Stierhof, I. Hindennach, and U. Henning. 1996. Characterization of the helper proteins for the assembly of tail fibers of coliphages T4 and lambda. J. Bacteriol. 178:6258–6265.
138. Hatfull, G. F., and G. J. Sarkis. 1993. DNA sequence, structure and gene expression of mycobacteriophage L5: a phage system for mycobacterial genetics. Mol. Microbiol. 7:395–405.
139. Hayden, M., M. B. Adams, and S. Casjens. 1985. Bacteriophage L: chromosome physical map and structural proteins. Virology 147:431–440.
140. Hendrix, R. 1971. Identification of proteins coded by phage lambda, pp. 355–370. *In* A. D. E. Hershey (ed.) The Bacteriophage Lambda. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
141. Hendrix, R. 1983. Progress since 1970, pp. 13–19. *In* R. Hendrix, J. Roberts, F. Stahl, and R. E. Weisberg (eds.) Lambda II. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
142. Hendrix, R., J. Roberts, F. Stahl, and R. E. Weisberg. 1983. Lambda II. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
143. Hendrix, R. W. 2002. Bacteriophages: evolution of the majority. Theor. Pop. Biol. 61:471–480.
144. Hendrix, R. W., and S. R. Casjens. 1975. Assembly of bacteriophage lambda heads: protein processing and its genetic control in petit lambda assembly. J. Mol. Biol. 91:187–199.
145. Hendrix, R. W., and S. R. Casjens. 1974. Protein cleavage in bacteriophage lambda tail assembly. Virology 61:156–159.
146. Hendrix, R. W., and S. R. Casjens. 1974. Protein fusion during the assembly of phage lambda heads. J. Supramol. Struct. 2:329–336.
147. Hendrix, R. W., and S. R. Casjens. 1974. Protein fusion: a novel reaction in bacteriophage lambda head assembly. Proc. Natl. Acad. Sci. USA 71:1451–1455.
148. Hendrix, R. W., and R. L. Duda. 1998. Bacteriophage HK97 head assembly: a protein ballet. Adv. Virus Res. 50:235–288.
149. Hendrix, R. W., and R. L. Duda. 1992. Bacteriophage lambda PaPa: not the mother of all lambda phages. Science 258:1145–1148.
150. Hendrix, R.W., J. G. Lawrence, G. F. Hatfull, and S. Casjens. 2000. The origins and ongoing evolution of viruses. Trends Microbiol. 8:504–508.
151. Hendrix, R. W., M. C. Smith, R. N. Burns, M. E. Ford, and G. F. Hatfull. 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. Proc. Natl. Acad. Sci. USA 96:2192–2197.
152. Henthorn, K. S., and D. I. Friedman. 1995. Identification of related genes in phages ϕ80 and P22 whose products are inhibitory for phage growth in *Escherichia coli* IHF mutants. J. Bacteriol. 177:3185–3190.
153. Herman, C., D. Thevenet, R. D'Ari, and P. Bouloc. 1997. The HflB protease of *Escherichia coli* degrades its inhibitor lambda cIII. J. Bacteriol. 179:358–363.
154. Hershey, A., and W. Dove. 1971. Introduction to lambda, pp. 3–11. *In* A. D. E. Hershey (ed.) The Bacteriophage Lambda. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
155. Hershey, A. D., and E. Burgi. 1965. Complementary structure of interacting sites at the ends of lambda DNA molecules. Proc. Natl. Acad. Sci. USA 53:325–330.
156. Hershey, A. D. E. 1971. The bacteriophage Lambda. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
157. Herskowitz, I., and D. Hagen. 1980. The lysis–lysogeny decision of phage lambda: explicit programming and responsiveness. Annu. Rev. Genet. 14:399–445.
158. Heymann, J. B., N. Cheng, W. W. Newcomb, B. L. Trus, J. C. Brown, and A. C. Steven. 2003. Dynamics of herpes simplex virus capsid maturation visualized by time-lapse cryo-electron microscopy. Nat. Struct. Biol. 10:334–341.
159. Highton, P. J., Y. Chang, and R. J. Myers. 1990. Evidence for the exchange of segments between genomes during the evolution of lambdoid bacteriophages. Mol. Microbiol. 4:1329–1340.
160. Ho, Y., and M. Rosenberg. 1982. Characterization of the phage lambda regulatory protein cII. Ann. Microbiol. (Paris) 133:215–218.
161. Ho, Y. S., M. E. Mahoney, D. L. Wulff, and M. Rosenberg. 1988. Identification of the DNA binding domain of the phage lambda cII transcriptional activator and the direct correlation of cII protein stability with its oligomeric forms. Genes Dev. 2:184–195.
162. Ho, Y. S., D. Pfarr, J. Strickler, and M. Rosenberg. 1992. Characterization of the transcription activator protein C1 of bacteriophage P22. J. Biol. Chem. 267:14388–14397.
163. Ho, Y. S., and M. Rosenberg. 1985. Characterization of a third, cII-dependent, coordinately activated promoter on phage lambda involved in lysogenic development. J. Biol. Chem. 260:11838-11844.
164. Ho, Y. S., D. L. Wulff, and M. Rosenberg. 1983. Bacteriophage lambda protein cII binds promoters on the opposite face of the DNA helix from RNA polymerase. Nature 304:703–708.
165. Hochschild, A., N. Irwin, and M. Ptashne. 1983. Repressor structure and the mechanism of positive control. Cell 32:319–325.
166. Hoffman, B., and M. Levine. 1975. Bacteriophage P22 virion protein which performs an essential early function. I. Analysis of *16*$^{-}$*ts* mutants. J. Virol. 16:1536–1546.
167. Hoffman, B., and M. Levine. 1975. Bacteriophage P22 virion protein which performs an essential early function. II. Characterization of the gene *16* function. J. Virol. 16:1547–1559.
168. Hohn, T., H. Flick, and B. Hohn. 1975. Petit lambda, a family of particles from coliphage lambda infected cells. J. Mol. Biol. 98:107–120.

169. Hong, J. S., G. R. Smith, and B. N. Ames. 1971. Adenosine 3′:5′-cyclic monophosphate concentration in the bacterial host regulates the viral decision between lysogeny and lysis. Proc. Natl. Acad. Sci. USA 68:2258–2262.
170. Huang, W. M., L. Joss, T. T. Hsieh, and S. Casjens. 2004. Protelomerase uses a topoisomerse IB/Y-recombinase type mechanism to generate DNA hairpin ends. J. Mol. Biol. 337: 77–92.
171. Imber, R., A. Tsugita, M. Wurtz, and T. Hohn. 1980. Outer surface protein of bacteriophage lambda. J. Mol. Biol. 139:277–295.
172. Israel, V. 1977. E proteins of bacteriophage P22. I. Identification and ejection from wild–type and defective particles. J. Virol. 23:91–97.
173. Israel, V. 1976. Role of the bacteriophage P22 tail in the early stages of infection. J. Virol. 18:361–364.
174. Israel, V., H. Rosen, and M. Levine. 1972. Binding of bacteriophage P22 tail parts to cells. J. Virol. 10:1152–1158.
175. Iwashita, S., and S. Kanegasaki. 1973. Smooth specific phage adsorption: endorhamnosidase activity of tail parts of P22. Biochem. Biophys. Res. Commun. 55:403–409.
176. Iyer, L. M., E. V. Koonin, and L. Aravind. 2002. Classification and evolutionary history of the single-strand annealing proteins, RecT, Redbeta, ERF and RAD52. BMC Genomics 3:8.
177. Jackson, E. N., D. A. Jackson, and R. J. Deans. 1978. *Eco*RI analysis of bacteriophage P22 DNA packaging. J. Mol. Biol. 118:365–388.
178. Jiang, W., Z. Li, Z. Zhang, M. L. Baker, P. E. Prevelige, Jr., and W. Chiu. 2003. Coat protein fold and maturation transition of bacteriophage P22 seen at subnanometer resolutions. Nat. Struct. Biol. 10:131–135.
179. Johnson, G., W. Widner, W. N. Xin, and M. Feiss. 1991. Interference with phage lambda development by the small subunit of the phage 21 terminase, gp1. J. Bacteriol. 173:2733–2738.
180. Jordan, E., L. Green, and H. Echols. 1973. Establishment of repression by bacteriophage lambda: lack of a direct regulatory effect of cyclic AMP. Virology 55:521–523.
181. Juhala, R. J., M. E. Ford, R. L. Duda, A. Youlton, G. F. Hatfull, and R. W. Hendrix. 2000. Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. J. Mol. Biol. 299:27–51.
182. Kaiser, A. D. 1957. Mutations in a temperate bacteriophage affecting its ability to lysogenize *Escherichia coli*. Virology 3:42–49.
183. Kaiser, A. D., and F. Jacob. 1957. Recombination between related bacteriophages and the genetic control of immunity and prophage localization. Virology 4:509–517.
184. Kaiser, D., and T. Masuda. 1973. In vitro assembly of bacteriophage lambda heads. Proc. Natl. Acad. Sci. USA 70:260–264.
185. Kaiser, D., M. Syvanen, and T. Masuda. 1975. DNA packaging steps in bacteriophage lambda head assembly. J. Mol. Biol. 91:175–186.
186. Kaiser, K. 1980. The origin of Q-independent derivatives of phage Lambda. Mol. Gen. Genet. 179:547–554.
187. Kaiser, K., and N. E. Murray. 1979. Physical characterisation of the ıRac prophageı in *E. coli* K12. Mol. Gen. Genet. 175:159–174.
188. Kanamaru, S., P. G. Leiman, V. A. Kostyuchenko, P. R. Chipman, V. V. Mesyanzhinov, F. Arisaka, and M. G. Rossmann. 2002. Structure of the cell-puncturing device of bacteriophage T4. Nature 415:553–557.
189. Kar, U. 1983. Structural proteins of bacteriophage lambda: purification, characterization and localization. PhD thesis, University of Pittsburgh, Pittsburgh, Pa.
190. Karakousis, G., N. Ye, Z. Li, S. K. Chiu, G. Reddy, and C. M. Radding. 1998. The beta protein of phage lambda binds preferentially to an intermediate in DNA renaturation. J. Mol. Biol. 276:721–731.
191. Karch, H., H. Schmidt, C. Janetzki-Mittmann, J. Scheef, and M. Kroger. 1999. Shiga toxins even when different are encoded at identical positions in the genomes of related temperate bacteriophages. Mol. Gen. Genet. 262:600–607.
192. Katsura, I. 1987. Determination of bacteriophage lambda tail length by a protein ruler. Nature 327:73–75.
193. Katsura, I. 1976. Morphogenesis of bacteriophage lambda tail. Polymorphism in the assembly of the major tail protein. J. Mol. Biol. 107:307–326.
194. Katsura, I. 1981. Structure and function of the major tail protein of bacteriophage lambda. Mutants having small major tail protein molecules in their virion. J. Mol. Biol. 146:493–512.
195. Katsura, I. 1983. Tail assembly and injection, pp. 331–346. *In* R. Hendrix, J. Roberts, F. W. Stahl, and R. Weisberg (eds.) Lambda II. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
196. Katsura, I., and R. W. Hendrix. 1984. Length determination in bacteriophage lambda tails. Cell 39:691–698.
197. Katsura, I., and P. W. Kuhl. 1975. Morphogenesis of the tail of bacteriophage lambda. II. In vitro formation and properties of phage particles with extra long tails. Virology 63:238–251.
198. Katsura, I., and P. W. Kuhl. 1975. Morphogenesis of the tail of bacteriophage lambda. III. Morphogenetic pathway. J. Mol. Biol. 91:257–273.
199. Katsura, I., and A. Tsugita. 1977. Purification and characterization of the major protein and the terminator protein of the bacteriophage lambda tail. Virology 76:129–145.
200. Kellenberger, E., and R. Edgar. 1971. Structure and assembly of phage particles, pp. 271–295. *In* A. D. Hershey (ed.) The Bacteriophge Lambda. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
201. Kellenberger, G., M. Zichichi, and J. Weigle. 1961. A mutation affecting the DNA content of bacteriophage λ and its lysogenizing properties. J. Mol. Biol. 3:399–408.
202. Kihara, A., Y. Akiyama, and K. Ito. 1997. Host regulation of lysogenic decision in bacteriophage lambda: transmembrane modulation of FtsH (HflB), the cII degrading protease, by HflKC (HflA). Proc. Natl. Acad. Sci. USA 94:5544–5549.

203. Kihara, A., Y. Akiyama, and K. Ito. 2001. Revisiting the lysogenization control of bacteriophage lambda. Identification and characterization of a new host component, HflD. J. Biol. Chem. 276:13695–13700.
204. Kim, H. C., J. G. Zhou, H. R. Wilson, G. Mogilnitskiy, D. L. Court, and M. E. Gottesman. 2003. Phage HK022 Nun protein represses translation of phage lambda N (transcription termination/translation repression). Proc. Natl. Acad. Sci. USA 100:5308–5312.
205. King, G., and N. E. Murray. 1995. Restriction alleviation and modification enhancement by the Rac prophage of *Escherichia coli* K-12. Mol. Microbiol. 16:769–777.
206. King, J., and S. Casjens. 1974. Catalytic head assembling protein in virus morphogenesis. Nature 251:112–119.
207. King, J., E. V. Lenk, and D. Botstein. 1973. Mechanism of head assembly and DNA encapsulation in *Salmonella* phage P22. II. Morphogenetic pathway. J. Mol. Biol. 80:697–731.
208. King, R. A., P. L. Madsen, and R. A. Weisberg. 2000. Constitutive expression of a transcription termination factor by a repressed prophage: promoters for transcribing the phage HK022 *nun* gene. J. Bacteriol. 182:456–462.
209. Kobiler, O., S. Koby, D. Teff, D. Court, and A. B. Oppenheim. 2002. The phage lambda CII transcriptional activator carries a C–terminal domain signaling for rapid proteolysis. Proc. Natl. Acad. Sci. USA 99:14964–14969.
210. Kochan, J., J. L. Carrascosa, and H. Murialdo. 1984. Bacteriophage lambda preconnectors. Purification and structure. J. Mol. Biol. 174:433–447.
211. Kornitzer, D., S. Altuvia, and A. B. Oppenheim. 1991. The activity of the CIII regulator of lambdoid bacteriophages resides within a 24-amino acid protein domain. Proc. Natl. Acad. Sci. USA 88:5217–5221.
212. Kovall, R., and B. W. Matthews. 1997. Toroidal structure of lambda-exonuclease. Science 277:1824–1827.
213. Krinke, L., M. Mahoney, and D. L. Wulff. 1991. The role of the OOP antisense RNA in coliphage lambda development. Mol. Microbiol. 5:1265–1272.
214. Krinke, L., and D. L. Wulff. 1990. RNase III-dependent hydrolysis of lambda *cII-O* gene mRNA mediated by lambda OOP antisense RNA. Genes Dev. 4:2223–2233.
215. Kropinski, A. M. 2000. Sequence of the genome of the temperate, serotype-converting, *Pseudomonas aeruginosa* bacteriophage D3. J. Bacteriol. 182:6066–6074.
216. Kusano, K., N. K. Takahashi, H. Yoshikura, and I. Kobayashi. 1994. Involvement of RecE exonuclease and RecT annealing protein in DNA double-strand break repair by homologous recombination. Gene 138:17–25.
217. Kwon, H. J., R. Tirumalai, A. Landy, and T. Ellenberger. 1997. Flexibility in DNA recombination: structure of the lambda integrase catalytic core. Science 276:126–131.
218. Landsmann, J., M. Kroger, and G. Hobom. 1982. The *rex* region of bacteriophage lambda: two genes under three-way control. Gene 20:11–24.
219. Lawrence, J. G., G. Hatfull, and R. Hendrix. 2002. The imbroglios of viral taxonomy: genetic exchange and the failings of phenetic approaches. J. Bacteriol. 184:4891–4905.
220. Learn, B. A., S. J. Um, L. Huang, and R. McMacken. 1997. Cryptic single-stranded-DNA binding activities of the phage lambda P and *Escherichia coli* DnaC replication initiation proteins facilitate the transfer of *E. coli* DnaB helicase onto DNA. Proc. Natl. Acad. Sci. USA 94:1154–1159.
221. Lederberg, E. 1951. Lysogenicity in *E. coli* K-12. Genetics 36:560.
222. Leffers, G. G., Jr., and S. Gottesman. 1998. Lambda Xis degradation in vivo by Lon and FtsH. J. Bacteriol. 180:1573–1577.
223. Leong, J. M., S. Nunes-Duby, A. Oser, C. F. Lesser, P. Youderian, M. M. Susskind, and A. Landy. 1984. Site-specific recombination systems of phages φ80 and P22: binding sites of integration host factor and recombination–induced mutations. Cold Spring Harb. Symp. Quant. Biol. 49:707–714.
224. Levin, M. E., R. W. Hendrix, and S. R. Casjens. 1993. A programmed translational frameshift is required for the synthesis of a bacteriophage λ tail assembly protein. J. Mol. Biol. 234:124–139.
225. Li, Z., G. Karakousis, S. K. Chiu, G. Reddy, and C. M. Radding. 1998. The beta protein of phage lambda promotes strand exchange. J. Mol. Biol. 276:733–744.
226. Liao, S. M., T. H. Wu, C. H. Chiang, M. M. Susskind, and W. R. McClure. 1987. Control of gene expression in bacteriophage P22 by a small antisense RNA. I. Characterization in vitro of the Psar promoter and the sar RNA transcript. Genes Dev. 1:197–203.
227. Lin, L., R. Bitner, and G. Edlin. 1977. Increased reproductive fitness of *Escherichia coli* lambda lysogens. *J. Virol.* 21:554–559.
228. Lindsey, D. F., C. Martinez, and J. R. Walker. 1992. Physical map location of the *Escherichia coli* attachment site for the P22 prophage (attP22). J. Bacteriol. 174:3834–3835.
229. Little, J. W. 1967. An exonuclease induced by bacteriophage lambda. II. Nature of the enzymatic reaction. J. Biol. Chem. 242:679–686.
230. Little, J. W. 1995. The SOS regulatory system, pp. 453–479. *In* E. Lynn and A. Lynch (eds.), Regulation of Gene Expression in *E. coli*. R. G. Landis, Georgetown, Tex.
231. Little, J. W., D. P. Shepley, and D. W. Wert. 1999. Robustness of a gene regulatory circuit. EMBO J. 18:4299–4307.
232. Lobocka, M. B., A. N. Svarchevsky, V. N. Rybchin, and M. B. Yarmolinsky. 1996. Characterization of the primary immunity region of the *Escherichia coli* linear plasmid prophage N15. J. Bacteriol. 178:2902–2910.
233. Loenen, W. A., and N. E. Murray. 1986. Modification enhancement by the restriction alleviation protein (Ral) of bacteriophage lambda. J. Mol. Biol. 190:11–22.
234. Mahajna, J., A. B. Oppenheim, A. Rattray, and M. Gottesman. 1986. Translation initiation of bacteriophage lambda gene *cII* requires integration host factor. J. Bacteriol. 165:167–174.
235. Mahdi, A. A., G. J. Sharples, T. N. Mandal, and R. G. Lloyd. 1996. Holliday junction resolvases encoded by homologous *rusA* genes in *Escherichia coli* K-12 and phage 82. J. Mol. Biol. 257:561–573.

236. Mallory, J. B., C. Alfano, and R. McMacken. 1990. Host virus interactions in the initiation of bacteriophage lambda DNA replication. Recruitment of *Escherichia coli* DnaB helicase by lambda P replication protein. J. Biol. Chem. 265:13297–13307.
237. Marr, M. T., S. A. Datwyler, C. F. Meares, and J. W. Roberts. 2001. Restructuring of an RNA polymerase holoenzyme elongation complex by lambdoid phage Q proteins. Proc. Natl. Acad. Sci. USA 98:8972–8978.
238. Marsic, N., S. Roje, I. Stojiljkovic, E. Salaj-Smic, and Z. Trgovcevic. 1993. In vivo studies on the interaction of RecBCD enzyme and lambda Gam protein. J. Bacteriol. 175:4738–4743.
239. Maxwell, K. L., A. A. Yee, C. H. Arrowsmith, M. Gold, and A. R. Davidson. 2002. The solution structure of the bacteriophage lambda head–tail joining protein, gpFII. J. Mol. Biol. 318:1395–1404.
240. Maxwell, K. L., A. A. Yee, V. Booth, C. H. Arrowsmith, M. Gold, and A. R. Davidson. 2001. The solution structure of bacteriophage lambda protein W, a small morphogenetic protein possessing a novel fold. J. Mol. Biol. 308:9–14.
241. McAdams, H. H., and L. Shapiro. 1995. Circuit simulation of genetic networks. Science 269:650–656.
242. McClelland, M., K. E. Sanderson, J. Spieth, S. W. Clifton, P. Latreille, L. Courtney, S. Porwollik, J. Ali, M. Dante, F. Du, S. Hou, D. Layman, S. Leonard, C. Nguyen, K. Scott, A. Holmes, N. Grewal, E. Mulvaney, E. Ryan, H. Sun, L. Florea, W. Miller, T. Stoneking, M. Nhan, R. Waterston, and R. K. Wilson. 2001. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. Nature 413:852–856.
243. McLenigan, M. P., O. I. Kulaeva, D. G. Ennis, A. S. Levine, and R. Woodgate. 1999. The bacteriophage P1 HumD protein is a functional homolog of the prokaryotic UmuD′-like proteins and facilitates SOS mutagenesis in *Escherichia coli*. J. Bacteriol. 181:7005–7013.
244. Mediavilla, J., S. Jain, J. Kriakov, M. E. Ford, R. L. Duda, W. R. Jacobs, Jr., R. W. Hendrix, and G. F. Hatfull. 2000. Genome organization and characterization of mycobacteriophage Bxb1. Mol. Microbiol. 38:955–970.
245. Mensa-Wilmot, K., R. Seaby, C. Alfano, M. C. Wold, B. Gomes, and R. McMacken. 1989. Reconstitution of a nine-protein system that initiates bacteriophage lambda DNA replication. J. Biol. Chem. 264:2853–2861.
246. Mikawa, Y., I. Maruyama, and S. Brenner. 1996. Surface display of proteins on bacteriophage lambda heads. J. Mol. Biol. 262:21–30.
247. Mmolawa, P. T., H. Schmieger, C. P. Tucker, and M. W. Heuzenroeder. 2003. Genomic structure of the *Salmonella enterica* serovar Typhimurium DT 64 bacteriophage ST64T: evidence for modular genetic architecture. J. Bacteriol. 185:3473–3475.
248. Moak, M., and I. Molineux. 2004. Peptidoglycan hydrolytic activities associated with bacteriophage virions. Mol. Microbiol. 51:1169–1183.
249. Morgan, G., G. Hatfull, S. Casjens, and R. Hendrix. 2002. Bacteriophage Mu genome sequence: analysis and comparison with Mu–like prophages in *Haemophilus, Neisseria* and *Deinococcus*. J. Mol. Biol. 317:337–359.
250. Mount, D., A. Harris, C. Fuerst, and L. Siminovitch. 1968. Mutation in bacteriophage lambda affecting particle morphogenesis. Virology 35:134–145.
251. Murialdo, H., and A. Becker. 1978. Head morphogenesis of complex double-stranded DNA bacteriophages. Bacteriol. Rev. 42:529–578.
252. Murialdo, H., and P. N. Ray. 1975. Model for arrangement of minor structural proteins in head of bacteriophage lambda. Nature 257:815–817.
253. Murialdo, H., and D. Tzamtzis. 1997. Mutations of the coat protein gene of bacteriophage lambda that overcome the necessity for the *FI* gene; the EFI domain. Mol. Microbiol. 24:341–353.
254. Murialdo, H., D. Tzamtzis, M. Berru, W. L. Fife, and A. Becker. 1997. Mutations in the terminase genes of bacteriophage lambda that bypass the necessity for *FI*. Mol. Microbiol. 24:937–952.
255. Murphy, K. C. 2000. Bacteriophage P22 Abc2 protein binds to RecC, increases the 5′ strand nicking activity of RecBCD, and together with lambda bet, promotes Chi-independent recombination. J. Mol. Biol. 296:385–401.
256. Murphy, K. C. 1998. Use of bacteriophage lambda recombination functions to promote gene replacement in *Escherichia coli*. J. Bacteriol. 180:2063–2071.
257. Neely, M. N., and D. I. Friedman. 1998. Arrangement and functional identification of genes in the regulatory region of lambdoid phage H-19B, a carrier of a Shiga–like toxin. Gene 223:105–113.
258. Neely, M. N., and D. I. Friedman. 2000. N-mediated transcription antitermination in lambdoid phage H-19B is characterized by alternative NUT RNA structures and a reduced requirement for host factors. Mol. Microbiol. 38:1074–1085.
259. Nickels, B. E., C. W. Roberts, H. Sun, J. W. Roberts, and A. Hochschild. 2002. The sigma^{70} subunit of RNA polymerase is contacted by the λ Q antiterminator during early elongation. *Mol. Cell* 10:611–622.
260. Novick, S. L., and J. D. Baldeschwieler. 1988. Fluorescence measurement of the kinetics of DNA injection by bacteriophage lambda into liposomes. Biochemistry 27:7919–7924.
261. Oberto, J., R. A. Weisberg, and M. E. Gottesman. 1989. Structure and function of the *nun* gene and the immunity region of the lambdoid phage HK022. J. Mol. Biol. 207:675–693.
262. Osawa, R., S. Iyoda, S. I. Nakayama, A. Wada, S. Yamai, and H. Watanabe. 2000. Genotypic variations of Shiga toxin-converting phages from enterohaemorrhagic *Escherichia coli* O157: H7 isolates. J. Med. Microbiol. 49:565–574.
263. Parma, D. H., M. Snyder, S. Sobolevski, M. Nawroz, E. Brody, and L. Gold. 1992. The Rex system of bacteriophage lambda: tolerance and altruistic cell death. Genes Dev. 6:497–510.
264. Pedulla, M. L., M. E. Ford, J. M. Houtz, T. Karthikeyan, C. Wadsworth, J. A. Lewis, D. Jacobs-Sera, J. Falbo, J. Gross, N. R. Pannunzio, W. Brucker, V. Kumar, J. Kandasamy, L. Keenan, S. Bardarov, J. Kriakov, J. G. Lawrence,

W. R. Jacobs, R. W. Hendrix, and G. F. Hatfull. 2003. Origins of highly mosaic mycobacteriophage genomes. Cell 113:171–182.

265. Pedulla, M. L., M. E. Ford, T. Karthikeyan, J. M. Houtz, R. W. Hendrix, G. F. Hatfull, A. R. Poteete, E. B. Gilcrease, D. A. Winn-Stapley, and S. R. Casjens. 2003. Corrected sequence of the bacteriophage P22 genome. J. Bacteriol. 185:1475–1477.

266. Perna, N. T., G. Plunkett, 3rd, V. Burland, B. Mau, J. D. Glasner, D. J. Rose, G. F. Mayhew, P. S. Evans, J. Gregor, H. A. Kirkpatrick, G. Posfai, J. Hackett, S. Klink, A. Boutin, Y. Shao, L. Miller, E. J. Grotbeck, N. W. Davis, A. Lim, E. T. Dimalanta, K. D. Potamousis, J. Apodaca, T. S. Anantharaman, J. Lin, G. Yen, D. C. Schwartz, R. A. Welch, and F. R. Blattner. 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature 409:529–533.

267. Perucchetti, R., W. Parris, A. Becker, and M. Gold. 1988. Late stages in bacteriophage lambda head morphogenesis: in vitro studies on the action of the bacteriophage lambda D-gene and W-gene products. Virology 165:103–114.

268. Plunkett, G., 3rd, and H. Echols. 1989. Retroregulation of the bacteriophage lambda *int* gene: limited secondary degradation of the RNase III-processed transcript. J. Bacteriol. 171:588-592.

269. Plunkett, G., 3rd, D. J. Rose, T. J. Durfee, and F. R. Blattner. 1999. Sequence of Shiga toxin 2 phage 933W from *Escherichia coli* O157:H7:Shiga toxin as a phage late–gene product. J. Bacteriol. 181:1767–1778.

270. Polissi, A., L. Goffin, and C. Georgopoulos. 1995. The *Escherichia coli* heat shock response and bacteriophage lambda development. FEMS Microbiol. Rev. 17:159–169.

271. Popa, M. P., T. A. McKelvey, J. Hempel, and R. W. Hendrix. 1991. Bacteriophage HK97 structure: wholesale covalent cross–linking between the major head shell subunits. J. Virol. 65:3227–3237.

272. Poteete, A. R. 1988. Bacteriophage P22, pp. 647–682. *In* R. Calendar (ed.) The Bacteriophages, vol. II. Plenum Press, New York.

273. Poteete, A. R., and A. C. Fenton. 1993. Efficient double-strand break-stimulated recombination promoted by the general recombination systems of phages lambda and P22. Genetics 134:1013–1021.

274. Poteete, A. R., and A. C. Fenton. 2000. Genetic requirements of phage lambda red-mediated gene replacement in *Escherichia coli* K-12. J. Bacteriol. 182:2336–2340.

275. Poteete, A. R., A. C. Fenton, and K. C. Murphy. 1988. Modulation of Escherichia coli RecBCD activity by the bacteriophage lambda Gam and P22 Abc functions. J. Bacteriol. 170:2012-2021.

276. Poteete, A. R., A. C. Fenton, and A. V. Semerjian. 1991. Bacteriophage P22 accessory recombination function. Virology 182:316–323.

277. Poteete, A. R., and M. Ptashne. 1982. Control of transcription by the bacteriophage P22 repressor. J. Mol. Biol. 157:21–48.

278. Poteete, A. R., R. T. Sauer, and R. W. Hendrix. 1983. Domain structure and quaternary organization of the bacteriophage P22 Erf protein. J. Mol. Biol. 171:401–418.

279. Potrykus, K., G. Wegrzyn, and V. J. Hernandez. 2002. Multiple mechanisms of transcription inhibition by ppGpp at the λP_R promoter. J. Biol. Chem. 277:43785–43791.

280. Prasad, B. V., P. E. Prevelige, E. Marietta, R. O. Chen, D. Thomas, J. King, and W. Chiu. 1993. Three-dimensional transformation of capsids associated with genome packaging in a bacterial virus. J. Mol. Biol. 231:65–74.

281. Ptashne, M. 1992. A Genetic Switch: phage λ and higher organisms, 2nd edn. Cell Press and Blackwell Scientific, Cambridge, Mass.

282. Radding, C. M., J. Rosenzweig, J. Richards, and E. Cassuto. 1971. Separation and characterization of exonuclease, β protein and a complex of both. J. Biol. Chem. 146:2510–2512.

283. Ranade, K., and A. R. Poteete. 1993. Superinfection exclusion (*sieB*) genes of bacteriophages P22 and lambda. J. Bacteriol. 175:4712–4718.

284. Ranade, K., and A. R. Poteete. 1993. A switch in translation mediated by an antisense RNA. Genes Dev. 7:1498–1507.

285. Rao, R. N., and C. V. Raj. 1973. *Salmonella typhimurium* mutants affecting establishment of lysogeny. Mol. Gen. Genet. 125:119–123.

286. Ravin, N., and V. Ravin. 1994. A mulitcopy plasmid based on a mini-replicon of the temperate bacteriophage N15. Mol. Gen. Mikrobiol. Virusol. 1:37–39 (in Russian).

287. Ravin, N., T. Strakhova, and V. Kuprianov. 2001. The protelomerase of the phage–plasmid N15 is respnonsible for its maintenannce in linear form. J. Mol. Biol. 312:899–906.

288. Ravin, N., A. Svarchevsky, and G. Deho. 1999. The antiimmunity system of phage-plasmid N15: identification of the antirepressor gene and its control by a small processed RNA. Mol. Microbiol. 34:980–994.

289. Ravin, N. V. 2003. Mechanisms of replication and telomere resolution of the linear plasmid prophage N15. FEMS Microbiol. Lett. 221:1–6.

290. Ravin, N. V., O. I. Doroshenko, and V. K. Ravin. 1998. COS-region of temperate phage N15. Mol. Gen. Mikrobiol. Virusol. 2:17–20 (in Russian).

291. Ravin, V., N. Ravin, S. Casjens, M. E. Ford, G. F. Hatfull, and R. W. Hendrix. 2000. Genomic sequence and analysis of the atypical temperate bacteriophage N15. J. Mol. Biol. 299:53–73.

292. Ravin, V., and M. G. Shulga. 1970. The evidence of extrachromosomal location phage prophage N15. Virology 40:800–805.

293. Ray, P., and H. Murialdo. 1975. The role of gene Nu3 in bacteriophage lambda head morphogenesis. Virology 64:247–263.

294. Ray, P., and M. Pearson. 1974. Evidence for post-transcriptional control of the morphogenetic genes of bacteriophage lambda. J. Mol. Biol. 85:163–175.

295. Ray, P., and M. Pearson. 1975. Functional inactivation of the bacteriophage lambda morphogenetic gene mRNA. Nature 253:637-650.
296. Recktenwald, J., and H. Schmidt. 2002. The nucleotide sequence of Shiga toxin (Stx) 2e–encoding phage ϕP27 is not related to other Stx phage genomes, but the modular genetic structure is conserved. Infect. Immun. 70:1896–1908.
297. Redfield, R. J., and A. Campbell. 1987. Structure of cryptic λ prophages. J. Mol. Biol. 198:393–404.
298. Reisinger, G. R., A. Rietsch, W. Lubitz, and U. Blasi. 1993. Lambda kil-mediated lysis requires the phage context. Virology 193:1033–1036.
299. Rennell, D., S. E. Bouvier, L. W. Hardy, and A. R. Poteete. 1991. Systematic mutation of bacteriophage T4 lysozyme. J. Mol. Biol. 222:67–88.
300. Revet, B., B. von Wilcken-Bergmann, H. Bessert, A. Barker, and B. Muller-Hill. 1999. Four dimers of lambda repressor bound to two suitably spaced pairs of lambda operators form octamers and DNA loops over large distances. Curr. Biol. 9:151–154.
301. Richardson, J., and J. Greenblatt. 1996. Control of RNA chain elongation and termination, pp. 822–848. *In* F. Neidhardt (ed.) *Escherichia coli* and *Salmonella* Cellular and Molecular Biology. ASM Press, Washington, D.C.
302. Riede, I., M. Degen, and U. Henning. 1985. The receptor specificity of bacteriophages can be determined by a tail fiber modifying protein. EMBO J. 4:2343–2346.
303. Roberts, C. W., and J. W. Roberts. 1996. Base-specific recognition of the nontemplate strand of promoter DNA by *E. coli* RNA polymerase. Cell 86:495–501.
304. Roberts, J., C. Roberts, S. Hilliker, and D. Botstein. 1976. Transcription termination and regulation in bacteriophages P22 and lambda, pp. 707–718. *In* R. Losick and M. Chamberlin (ed.) RNA Polymerase. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
305. Roberts, J. W., and R. Devoret. 1983. Lysogenic induction, pp. 123–144. *In* R. W. Hendrix, J. Roberts, F. Stahl, and R. Weisberg (ed.) Lambda II. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
306. Roberts, J. W., W. Yarnell, E. Bartlett, J. Guo, M. Marr, D. C. Ko, H. Sun, and C. W. Roberts. 1998. Antitermination by bacteriophage lambda Q protein. Cold Spring Harb. Symp. Quant. Biol. 63:319–325.
307. Roessner, C. A., and G. M. Ihler. 1987. Sequence of amino acids in lamB responsible for spontaneous ejection of bacteriophage lambda DNA. J. Mol. Biol. 195:963–966.
308. Roessner, C. A., D. K. Struck, and G. M. Ihler. 1983. Injection of DNA into liposomes by bacteriophage lambda. J. Biol. Chem. 258:643–648.
309. Rolfe, B., J. Schell, A. Becker, J. Heip, K. Onodera, and E. Schell-Frederick. 1973. A colicin-tolerant mutant of *Escherichia coli* with reduced levels of cyclic AMP and a strong bias towards lambda lysogeny. Mol. Gen. Genet. 120:1–16.
310. Saad, A., Z. H. Zhou, J. Jakana, W. Chiu, and F. J. Rixon. 1999. Roles of triplex and scaffolding proteins in herpes simplex virus type 1 capsid formation suggested by structures of recombinant particles. J. Virol. 73:6821–6830.
311. Sampson, L. L., R. W. Hendrix, W. M. Huang, and S. R. Casjens. 1988. Translation initiation controls the relative rates of expression of the bacteriophage lambda late genes. Proc. Natl. Acad. Sci. USA 85:5439–5443.
312. Sandmeier, H., S. Iida, and W. Arber. 1992. DNA inversion regions Min of plasmid p15B and Cin of bacteriophage P1: evolution of bacteriophage tail fiber genes. J. Bacteriol. 174:3936–3944.
313. Sandt, C. H., J. E. Hopper, and C. W. Hill. 2002. Activation of prophage eib genes for immunoglobulin-binding proteins by genes from the IbrAB genetic island of *Escherichia coli* ECOR-9. J. Bacteriol. 184:3640–3648.
314. Sanger, F., A. R. Coulson, G. F. Hong, D. F. Hill, and G. B. Petersen. 1982. Nucleotide sequence of bacteriophage lambda DNA. J. Mol. Biol. 162:729–773.
315. Sauer, R. T., W. Krovatin, J. DeAnda, P. Youderian, and M. M. Susskind. 1983. Primary structure of the immI immunity region of bacteriophage P22. J. Mol. Biol. 168:699–713.
316. Sauer, R. T., H. C. Nelson, K. Hehir, M. H. Hecht, F. S. Gimble, J. DeAnda, and A. R. Poteete. 1983. The lambda and P22 phage repressors. J. Biomol. Struct. Dyn. 1:1011–1022.
317. Scandella, D., and W. Arber. 1976. Phage lambda DNA injection into *Escherichia coli pel*$^-$ mutants is restored by mutations in phage genes *V* or *H*. Virology 69:206–215.
318. Schaefer, K. L., and W. R. McClure. 1997. Antisense RNA control of gene expression in bacteriophage P22. I. Structures of sar RNA and its target, *ant* mRNA. *RNA* 3:141–156.
319. Scharpf, M., H. Sticht, K. Schweimer, M. Boehm, S. Hoffmann, and P. Rosch. 2000. Antitermination in bacteriophage lambda. The structure of the N36 peptide-boxB RNA complex. Eur. J. Biochem. 267:2397–2408.
320. Schicklmaier, P., E. Moser, T. Wieland, W. Rabsch, and H. Schmieger. 1998. A comparative study on the frequency of prophages among natural isolates of *Salmonella* and *Escherichia coli* with emphasis on generalized transducers. Antonie van Leeuwenhoek 73:49–54.
321. Schmieger, H. 1999. Molecular survey of the *Salmonella* phage typing system of Anderson. J. Bacteriol. 181:1630–1635.
322. Sergueev, K., D. Court, L. Reaves, and S. Austin. 2002. *E. coli* cell-cycle regulation by bacteriophage lambda. J. Mol. Biol. 324:297–307.
323. Shaw, J. E., and H. Murialdo. 1980. Morphogenetic genes *C* and *Nu3* overlap in bacteriophage lambda. Nature 283:30–35.
324. Shimatake, H., and M. Rosenberg. 1981. Purified lambda regulatory protein cII positively activates promoters for lysogenic development. Nature 292:128–132.
325. Shotland, Y., A. Shifrin, T. Ziv, D. Teff, S. Koby, O. Kobiler, and A. B. Oppenheim. 2000. Proteolysis of bacteriophage lambda CII by *Escherichia coli* FtsH (HflB). J. Bacteriol. 182:3111–3116.
326. Siegele, D. A., S. Frackman, J. Sippy, T. Momany, T. M. Howard, K. Tilly, C. Georgopoulos, and M. Feiss. 1983. The head genes of bacteriophage 21. Virology 129:484–489.

327. Simon, M., R. W. Davis, and N. Davidson. 1971. Heteroduplexs of DNA moleculaes of lambdoid phages: physical mapping of their base sequence relationships by electron microscopy, pp. 313–328. *In* A. D. Hershey (ed.) The Bacteriophage Lambda. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
328. Simpson, A. A., Y. Tao, P. G. Leiman, M. O. Badasso, Y. He, P. J. Jardine, N. H. Olson, M. C. Morais, S. Grimes, D. L. Anderson, T. S. Baker, and M. G. Rossmann. 2000. Structure of the bacteriophage ϕ29 DNA packaging motor. Nature 408:745–750.
329. Slavcev, R. A., and S. Hayes. 2003. Stationary phase-like properties of the bacteriophage lambda Rex exclusion phenotype. Mol. Genet. Genomics 269:40–48.
330. Slominska, M., G. Konopa, J. Ostrowska, B. Kedzierska, G. Wegrzyn, and A. Wegrzyn. 2003. SeqA-mediated stimulation of a promoter activity by facilitating functions of a transcription activator. Mol. Microbiol. 47:1669–1679.
331. Slominska, M., P. Neubauer, and G. Wegrzyn. 1999. Regulation of bacteriophage lambda development by guanosine 5′-diphosphate-3′-diphosphate. Virology 262:431–441.
332. Smith, G. R. 1983. General recombination, pp. 175-209. *In* R. W. Hendrix, J. W. Roberts, F. W. Stahl, and R. A. Weisberg (eds.) Lambda II. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
333. Smith, M. P., and M. Feiss. 1993. Sequence analysis of the phage 21 genes for prohead assembly and head completion. Gene 126:1–7.
334. Snyder, L. 1995. Phage-exclusion enzymes: a bonanza of biochemical and cell biology reagents? Mol. Microbiol. 15:415–420.
335. Stahl, F. W. 1998. Recombination in phage lambda: one geneticist's historical perspective. Gene 223:95-102.
336. Steinbacher, S., U. Baxa, S. Miller, A. Weintraub, R. Seckler, and R. Huber. 1996. Crystal structure of phage P22 tailspike protein complexed with *Salmonella* sp. O-antigen receptors. Proc. Natl. Acad. Sci. USA 93:10584–10588.
337. Steinbacher, S., S. Miller, U. Baxa, N. Budisa, A. Weintraub, R. Seckler, and R. Huber. 1997. Phage P22 tailspike protein: crystal structure of the head-binding domain at 2.3 Å, fully refined structure of the endorhamnosidase at 1.56 Å resolution, and the molecular basis of O-antigen recognition and cleavage. J. Mol. Biol. 267:865–880.
338. Steinbacher, S., R. Seckler, S. Miller, B. Steipe, R. Huber, and P. Reinemer. 1994. Crystal structure of P22 tailspike protein: interdigitated subunits in a thermostable trimer. Science 265:383–386.
339. Sternberg, N., and R. Hoess. 1995. Display of peptides and proteins on the surface of bacteriophage lambda. Proc. Natl. Acad. Sci. USA 92:1609–1613.
340. Sternberg, N., and R. Weisberg. 1977. Packaging of coliphage lambda DNA. II. The role of the gene D protein. J. Mol. Biol. 117:733–759.
341. Susskind, M. M., and D. Botstein. 1975. Mechanism of action of *Salmonella* phage P22 antirepressor. J. Mol. Biol. 98:413–424.
342. Susskind, M. M., and D. Botstein. 1978. Molecular genetics of bacteriophage P22. Microbiol. Rev. 42:385–413.
343. Susskind, M. M., D. Botstein, and A. Wright. 1974. Superinfection exclusion by P22 prophage in lysogens of *Salmonella typhimurium*. III. Failure of superinfecting phage DNA to enter *sieA*$^+$ lysogens. Virology 62:350–366.
344. Susskind, M. M., A. Wright, and D. Botstein. 1974. Superinfection exclusion by P22 prophage in lysogens of *Salmonella typhimurium*. IV. Genetics and physiology of *sieB* exclusion. Virology 62:367–384.
345. Szalewska-Palasz, A., A. Wegrzyn, M. Obuchowski, R. Pawlowski, K. Bielawski, M. Thomas, and G. Wegrzyn. 1996. Drastically decreased transcription from CII-activated promoters is responsible for impaired lysogenization of the *Escherichia coli* rpoA341 mutant by bacteriophage lambda. FEMS Microbiol. Lett. 144:21–27.
346. Szalewska-Palasz, A., B. Wrobel, and G. Wegrzyn. 1998. Rapid degradation of polyadenylated oop RNA. FEBS Lett. 432:70-72.
347. Tarkowski, T. A., D. Mooney, L. C. Thomason, and F. W. Stahl. 2002. Gene products encoded in the *ninR* region of phage lambda participate in Red-mediated recombination. Genes Cells 7:351–363.
348. Taylor, A., S. Kedzierska, and A. Wawrzynow. 1996. Bacteriophage lambda lysis gene product modified and inserted into *Escherichia coli* outer membrane: Rz1 lipoprotein. Microb. Drug Resist. 2:147–153.
349. Taylor, K., and G. Wegrzyn. 1995. Replication of coliphage lambda DNA. FEMS Microbiol. Rev. 17:109–119.
350. Thomas, J. O. 1974. Chemical linkage of the tail to the right-hand end of bacteriophage lambda DNA. J. Mol. Biol. 87:1–9.
351. Toothman, P., and I. Herskowitz. 1980. Rex-dependent exclusion of lambdoid phages. I. Prophage requirements for exclusion. Virology 102:133–146.
352. Toothman, P., and I. Herskowitz. 1980. Rex-dependent exclusion of lambdoid phages. II. Determinants of sensitivity to exclusion. Virology 102:147–160.
353. Tsui, L., and R. W. Hendrix. 1980. Head–tail connector of bacteriophage lambda. J. Mol. Biol. 142:419–438.
354. Tsui, L. C., and R. W. Hendrix. 1983. Proteolytic processing of phage lambda tail protein gpH: timing of the cleavage. Virology 125:257-264.
355. Tsurimoto, T., and K. Matsubara. 1981. Purified bacteriophage lambda O protein binds to four repeating sequences at the lambda replication origin. Nucleic Acids Res. 9:1789–1799.
356. Tye, B. K., and D. Botstein. 1974. P22 morphogenesis. II. Mechanism of DNA encapsulation. J. Supramol. Struct. 2:225–238.
357. Unkmeir, A., and H. Schmidt. 2000. Structural analysis of phage–borne *stx* genes and their flanking sequences in Shiga toxin-producing *Escherichia coli* and *Shigella dysenteriae* type 1 strains. Infect. Immun. 68:4856–4864.
358. Vaca Pacheco, S., O. Garcia Gonzalez, and G. L. Paniagua Contreras. 1997. The *lom* gene of bacteriophage lambda is involved in *Escherichia coli* K12 adhesion to human buccal epithelial cells. FEMS Microbiol. Lett. 156:129–132.

359. van der Wilk, F., A. M. Dullemans, M. Verbeek, and J. F. van den Heuvel. 1999. Isolation and characterization of APSE-1, a bacteriophage infecting the secondary endosymbiont of *Acyrthosiphon pisum.* Virology 262:104–113.
360. Vander Byl, C., and A. M. Kropinski. 2000. Sequence of the genome of *Salmonella* bacteriophage P22. J. Bacteriol. 182:6472–6481.
361. Vershon, A. K., P. Youderian, M. M. Susskind, and R. T. Sauer. 1985. The bacteriophage P22 arc and mnt repressors. Overproduction, purification, and properties. J. Biol. Chem. 260:12124–12129.
362. Voegtli, W. C., D. J. White, N. J. Reiter, F. Rusnak, and A. C. Rosenzweig. 2000. Structure of the bacteriophage lambda Ser/Thr protein phosphatase with sulfate ion bound in two coordination modes. Biochemistry 39:15365–15374.
363. Vostrov, A., O. Vostrukhina, A. Svarchevsky, and V. Rybchin. 1996. Proteins responsible for lysogenic conversion caused by coliphages N15 and ϕ80 are highly homologous. J. Bacteriol. 178:1484–1486.
364. Wagner, P. L., J. Livny, M. N. Neely, D. W. Acheson, D. I. Friedman, and M. K. Waldor. 2002. Bacteriophage control of Shiga toxin 1 production and release by *Escherichia coli.* Molec. Microbiol. 44:957–970.
365. Wagner, P. L., M. N. Neely, X. Zhang, D. W. Acheson, M. K. Waldor, and D. I. Friedman. 2001. Role for a phage promoter in Shiga toxin 2 expression from a pathogenic *Escherichia coli* strain. J. Bacteriol. 183:2081–2085.
366. Wagner, P. L., and M. K. Waldor. 2002. Bacteriophage control of bacterial virulence. Infect. Immun. 70:3985–3993.
367. Walker, J. E., A. D. Auffret, A. Carne, A. Gurnett, P. Hanisch, D. Hill, and M. Saraste. 1982. Solid–phase sequence analysis of polypeptides eluted from polyacrylamide gels. An aid to interpretation of DNA sequences exemplified by the *Escherichia coli unc* operon and bacteriophage lambda. Eur. J. Biochem. 123:253–260.
368. Wang, I. N., D. L. Smith, and R. Young. 2000. Holins: the protein clocks of bacteriophage infections. Annu. Rev. Microbiol. 54:799–825.
369. Wang, J., M. Hofnung, and A. Charbit. 2000. The C-terminal portion of the tail fiber protein of bacteriophage lambda is responsible for binding to LamB, its receptor at the surface of *Escherichia coli* K-12. J. Bacteriol. 182:508–512.
370. Weaver, L. H., D. Rennell, A. R. Poteete, and B. W. Mathews. 1985. Structure of phage P22 gene *19* lysozyme inferred from its homology with phage T4 lysozyme. Implications for lysozyme evolution. J. Mol. Biol. 184:739–741.
371. Weigle, J. 1966. Assembly of phage lambda in vitro. Proc. Natl. Acad. Sci. USA 55:1462–1466.
372. Weigle, J. 1968. Studies on head–tail union in bacteriophage lambda. J. Mol. Biol. 33:483–489.
373. Weisberg, R. 1996. Specialized transduction, pp. 2442–2448. *In* F. Neidhardt (ed.) *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology, vol. 2. ASM Press, Washington, D.C.
374. Weisberg, R., and M. Gottesman. 1971. The stability of Int and Xis functions, pp. 489–500. *In* A. D. Hershey (ed.) The Bacteriophage Lambda. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
375. Weisberg, R., and A. Landy. 1983. Site-specific recombination, pp. 211–250. *In* R. Hendrix, J. Roberts, F. W. Stahl, and R. Weisberg (eds.) Lambda II. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
376. Weisberg, R. A., and M. E. Gottesman. 1999. Processive antitermination. J. Bacteriol. 181:359–367.
377. Werts, C., V. Michel, M. Hofnung, and A. Charbit. 1994. Adsorption of bacteriophage lambda on the LamB protein of *Escherichia coli* K-12: point mutations in gene *J* of lambda responsible for extended host range. J. Bacteriol. 176:941–947.
378. Whitman, W. B., D. C. Coleman, and W. J. Wiebe. 1998. Prokaryotes: the unseen majority. Proc. Natl. Acad. Sci. USA 95:6578–6583.
379. Wickner, S. 1984. DNA-dependent ATPase activity associated with phage P22 gene *12* protein. J. Biol. Chem. 259:14038–14043.
380. Wickner, S. 1984. Oligonucleotide synthesis by *Escherichia coli dnaG* primase in conjunction with phage P22 gene *12* protein. J. Biol. Chem. 259:14044–14047.
381. Wickner, S. H., and K. Zahn. 1986. Characterization of the DNA binding domain of bacteriophage lambda O protein. J. Biol. Chem. 261:7537–7543.
382. Wikoff, W. R., L. Liljas, R. L. Duda, H. Tsuruta, R. W. Hendrix, and J. E. Johnson. 2000. Topologically linked protein rings in the bacteriophage HK97 capsid. Science 289:2129–2133.
383. Williams, N., D. K. Fox, C. Shea, and S. Roseman. 1986. Pel, the protein that permits lambda DNA penetration of *Escherichia coli*, is encoded by a gene in *ptsM* and is required for mannose utilization by the phosphotransferase system. Proc. Natl. Acad. Sci. USA 83:8934–8938.
384. Williams, R., and K. Richards. 1974. Capsid structure in bacteriophage lambda. J. Mol. Biol. 88:547–555.
385. Willis, D. K., L. H. Satin, and A. J. Clark. 1985. Mutation-dependent suppression of recB21 recC22 by a region cloned from the Rac prophage of *Escherichia coli* K-12. J. Bacteriol. 162:1166–1172.
386. Wolberger, C., Y. C. Dong, M. Ptashne, and S. C. Harrison. 1988. Structure of a phage 434 Cro/DNA complex. Nature 335:789–795.
387. Wommack, K. E., and R. R. Colwell. 2000. Virioplankton: viruses in aquatic ecosystems. Microbiol. Mol. Biol. Rev. 64:69–114.
388. Wood, W. B. 1980. Bacteriophage T4 morphogenesis as a model for assembly of subcellular structure. Q. Rev. Biol. 55:353–367.
389. Woods, D. E., J. A. Jeddeloh, D. L. Fritz, and D. DeShazer. 2002. *Burkholderia thailandensis* E125 harbors a temperate bacteriophage specific for *Burkholderia mallei.* J. Bacteriol. 184:4003–4017.
390. Wrobel, B., A. Herman-Antosiewicz, S. Szalewska-Palasz, and G. Wegrzyn. 1998. Polyadenylation of oop RNA in the regulation of bacteriophage lambda development. Gene 212:57-65.
391. Wrobel, B., and G. Wegrzyn. 2002. Evolution of lambdoid replication modules. Virus Genes 24:163–171.

392. Wu, H., L. Sampson, R. Parr, and S. Casjens. 2002. The DNA site utilized by bacteriophage P22 for initiation of DNA packaging. Mol. Microbiol. 45:1631–1646.
393. Wu, R., and A. D. Kaiser. 1968. Structure and base sequence in the cohesive ends of phage lambda. J. Mol. Biol. 57:491–499.
394. Wu, T. H., S. M. Liao, W. R. McClure, and M. M. Susskind. 1987. Control of gene expression in bacteriophage P22 by a small antisense RNA. II. Characterization of mutants defective in repression. Genes Dev. 1:204–212.
395. Wulff, D., and M. Rosenberg. 1983. Establishment of repressor synthesis, pp. 53–73. *In* R. W. Hendrix, J. Roberts, F. Stahl, and R. Weisberg (eds.) Lambda II. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
396. Wulff, D. L., Y. S. Ho, S. Powers, and M. Rosenberg. 1993. The *int* genes of bacteriophages P22 and lambda are regulated by different mechanisms. Molec. Microbiol. 9:261–271.
397. Wulff, D. L., and M. E. Mahoney. 1987. Cross-specificities between cII-like proteins and pRE-like promoters of lambdoid bacteriophages. Genetics 115:597–604.
398. Xin, W. N., and M. Feiss. 1988. The interaction of *Escherichia coli* integration host factor with the cohesive end sites of phages lambda and 21. Nucleic Acids Res. 16:2015–2030.
399. Xu, J. 2000. A conserved frameshift strategy in dsDNA long tailed phages. PhD thesis, University of Pittsburgh, Pittsburgh, Pa.
400. Yang, F., P. Forrer, Z. Dauter, J. F. Conway, N. Cheng, M. E. Cerritelli, A. C. Steven, A. Pluckthun, and A. Wlodawer. 2000. Novel fold and capsid-binding properties of the lambda-phage display platform protein gpD. Nat. Struct. Biol. 7:230-237.
401. Yang, X. J., J. A. Goliger, and J. W. Roberts. 1989. Specificity and mechanism of antitermination by Q proteins of bacteriophages lambda and 82. J. Mol. Biol. 210:453–460.
402. Yarmolinsky, M. B. 1971. Making and joining DNA ends, pp. 97-111. *In* A. D. Hershey (ed.) The Bacteriophage Lambda. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
403. Yochem, J., H. Uchida, M. Sunshine, H. Saito, C. P. Georgopoulos, and M. Feiss. 1978. Genetic analysis of two genes, *dnaJ* and *dnaK*, necessary for *Escherichia coli* and bacteriophage lambda DNA replication. Mol. Gen. Genet. 164:9–14.
404. Yokota, T., and T. Kasuga. 1972. Requirement of adenosine 3′,5′-cyclic phosphate for formation of the phage lambda receptor in *Escherichia coli*. J. Bacteriol. 109:1304–1306.
405. Yokoyama, K., K. Makino, Y. Kubota, C. H. Yutsudo, S. Kimura, K. Kurokawa, K. Ishii, M. Hattori, I. Tatsuno, H. Abe, T. Iida, K. Yamamoto, M. Onishi, T. Hayashi, T. Yasunaga, T. Honda, C. Sasakawa, and H. Shinagawa. 1999. Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic *Escherichia coli* O157:H7 derived from the Sakai outbreak. Genes Genet. Syst. 74:227–239.
406. Young, R. 2002. Bacteriophage holins: deadly diversity. J. Mol. Microbiol. Biotechnol. 4:21–36.
407. Zhang, N., and R. Young. 1999. Complementation and characterization of the nested Rz and Rz1 reading frames in the genome of bacteriophage lambda. Mol. Gen. Genet. 262:659–667.
408. Zylicz, M., D. Ang, K. Liberek, and C. Georgopoulos. 1989. Initiation of lambda DNA replication with purified host- and bacteriophage-encoded proteins: the role of the dnaK, dnaJ and grpE heat shock proteins. EMBO J. 8:1601–1608.
409. Zylicz, M., D. Ang, K. Liberek, T. Yamamoto, and C. Georgopoulos. 1988. Initiation of lambda DNA replication reconstituted with purified lambda and *Escherichia coli* replication proteins. Biochim. Biophys. Acta 951:344–350.
410. Zylicz, M., K. Liberek, A. Wawrzynow, and C. Georgopoulos. 1998. Formation of the preprimosome protects lambda O from RNA transcription-dependent proteolysis by ClpP/ClpX. Proc. Natl. Acad. Sci. USA 95:15259–15263.

28

N15: The Linear Plasmid Prophage

NIKOLAI V. RAVIN

The temperate bacteriophage N15 is unique among *Escherichia coli* phages in that its prophage is not integrated into the bacterial chromosome but instead is a linear plasmid molecule with covalently closed ends. Linear plasmids are very unusual in the Enterobacteria and only two others have been reported: phiKO2 in *Klebsiella oxytoca* (4, 43) and pY54 in *Yersinia enterocolitica* (16). However, such replicons are common in the spirochete genus, *Borrelia* (1). Here I will summarize the most relevant work on phage N15 with a special emphasis on the mechanism of replication, telomere formation, and control of lysogeny.

Bacteriophage N15 was isolated by Victor Ravin in 1964 and was initially studied in Moscow, Russia (13, 33–36, 38, 39). N15 belongs to the lambdoid phage family, as suggested on the basis of cross-hybridization of their DNAs (39), and is similar to phage λ with respect to the length of the genome, morphology of phage particles and plaques, burst size, latent period, and lysogenization frequency (34) (for more on lambdoid phages, see chapter 27). As was shown by Victor Ravin and coworkers in 1967–1970, an unusual feature of phage N15 is that its prophage locates extrachromosomally. Three lines of evidence led to this conclusion. First, it was observed that lysogenic bacteria lose the prophage at nonpermissive temperatures if they have temperature-sensitive mutations in "early" genes (33, 38). Second, results of bacterial crosses show an absence of the prophage in the bacterial chromosome (35). Third, total DNA isolated from lysogenic cultures can be separated into bacterial and phage fractions by sucrose gradient centrifugation (39).

The next important step in the investigation of N15 biology was performed by Valentin Rybchin and colleagues (Saint Petersburg, Russia), who showed that the N15 prophage is a linear plasmid with covalently closed ends. Linearity of the N15 prophage resulted from the physical mapping of phage and plasmid prophage DNAs, which showed (i) that both DNAs are linear molecules 46.3 kb long, and (ii) that the two maps are circularly permuted (45). The ends of the N15 prophage (telomeres), designated *telL* and *telR*, are covalently closed hairpins in which one strand simply turns around and becomes the other strand (22, 23, 45). Inference of the hairpin nature of the phage N15 ends results from two observations: (i) terminal, but not internal, restriction fragments of N15 plasmid DNA renature rapidly after heat denaturation and quick cooling; (ii) treatment of plasmid DNA prior to restriction enzyme digestion with the S1 nuclease, which is known to cut single-stranded DNA, abolishes the rapid renaturation effect (45). Mature N15 phage DNA, like phage λ DNA, has 12 bp single-stranded cohesive ends (26), named *cosL* and *cosR*.

The above data suggest the following mechanism of conversion of phage DNA to the prophage plasmid (22). After infection of an *E. coli* cell, the phage DNA becomes circularized via its cohesive termini. A special phage-encoded enzyme, protelomerase (prokaryotic telomerase), then introduces a staggered nick in the *telRL* region which contains a large palindromic sequence. Annealing of self-complementary single-stranded ends and formation of phosphodiester bonds results in creation of hairpin structures at each end (figure 28-1).

Overall Organization of the Genome

The nucleotide sequence of the N15 genome has been recently determined (37). The genome contains 46,363 bp, of which about one half are similar to the bacteriophage λ sequence (figure 28-2). This sequence similarity maps predominantly to the left arm of the phage genome (figure 28-2), which contains the structural genes for the proteins required for virion head and tail assembly. Genes *1* through *21* from phage N15 display a one-to-one correspondence to the phage λ genes *A* through *J*. There is as much as 90% identity between the amino acid sequences of the phage N15 and phage λ head gene products. Some parts of this region of phage N15 are more closely related to other, non-λ lambdoid phages. From gene *17* (the λ tail-assembly gene *M* analog) to gene *25*, except for gene *24*, phage N15 matches lambdoid phages HK97 and HK022 better than λ.

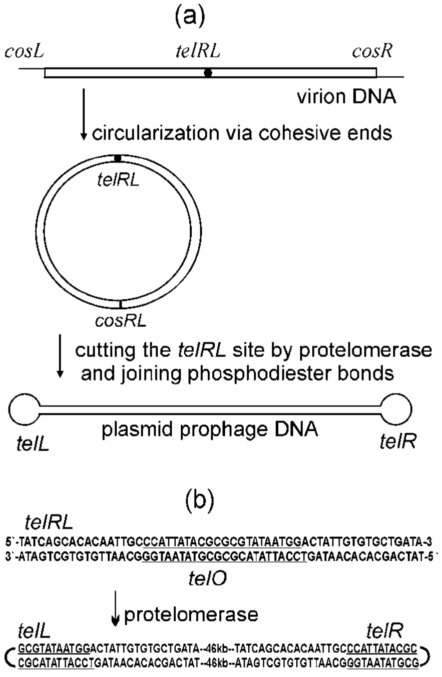

(a)
cosL
telRL
cosR
virion DNA
circularization via cohesive ends
telRL
cosRL
cutting the telRL site by protelomerase
and joining phosphodiester bonds
plasmid prophage DNA
telL
telR
(b)
telRL
5`-TATCAGCACACAATTGCCCATTATACGCGCGTATAATGGACTATTGTGTGCTGATA-3
3`-ATAGTCGTGTGTTAACGGGTAATATGCGCGCATATTACCTGATAACACACGACTAT-5`
telO
protelomerase
telL
telR
GCGTATAATGGACTATTGTGTGCTGATA--46kb--TATCAGCACACAATTGCCCATTATACGC
CGCATATTACCTGATAACACACGACTAT--46kb--ATAGTCGTGTGTTAACGGGTAATATGCG

Figure 28-1 Phage N15 chromosome linearization. A: Mechanism of conversion of phage DNA into linear plasmid. *cosL*, *cosR*, single-stranded cohesive ends; *cosRL*, represents the *cos* site-after annealing and ligation of cohesive ends; *telRL*, uncut target site of protelomerase; *telL* and *telR*, left and right hairpin ends of the prophage created by protelomerase. B: Sequences of *telRL* site and hairpin ends of the prophage. The central 22 bp ideal palindrome *telO* is underlined.

The above observation that phage N15 carries λ-like head- and tail-protein genes correlates well with the observation that N15 virion morphology is similar to that of lambdoid phages (37).

The N15 gene *24* is the homolog and functional analog of the *cor* gene of phage φ80 (49) and is responsible for inability of N15 lysogens to adsorb bacteriophages N15, T1, and φ80 (36). To the right of the just-described block of morphogenic genes there is gene *26*, a homolog of the *E.coli umuD* gene, which is involved in error-prone DNA damage repair. The next two genes in the left half of the genome are homolog of the *sopA* and *sopB* genes of the F plasmid and determine the segregation stability of the N15 prophage (see below).

The division between the left and right arms of the N15 genome is determined by the site (*telRL*) at which phage DNA is cut by protelomerase to make the linear plasmid prophage (figure 28-1). Contrary to the left arm, only 10 of the 35 N15 right-arm genes have homologs in lambdoid phages. Among these lambdoid-like genes are genes *38*, *39* and *40*, which are homolog to genes *cB*, *cro* and *Q*, respectively and are responsible for control of lysogeny (see below). Genes *53*, *54*, *55*, and *55.1* are thought to encode lysis functions and also have homologs in the lambdoid phage family (37). The three operons located in the right arm that are specific to phage N15 reflect its unusual lifestyle: the protelomerase gene (*29*), which is located rightward of *telRL*, the antirepressor operon including genes *30–32*; and the replication region comprising genes *33–37*, which are supposed to be cotranscribed from the promoter controlled by the CB repressor. Detailed description of other

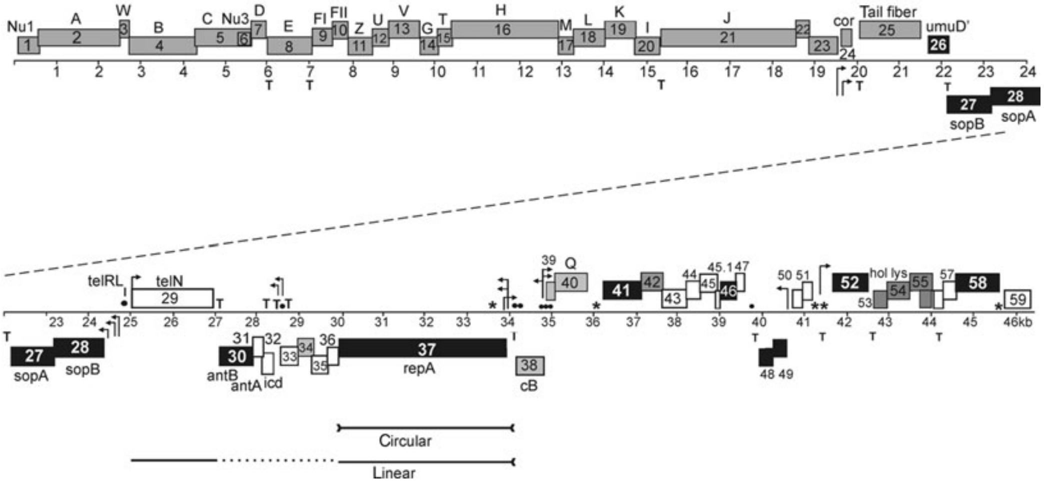

Nu1
A
W
B
C
Nu3
D
E
FI
FII
Z
U
V
G
T
H
M
L
K
I
J
cor
Tail fiber
umuD'
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
27 28
sopB
sopA
telRL
telN
29
T
39
Q
40
41 42 43 44 45 45.1 46 47
50 51
52
hol lys
53 54 55 57 58 59
23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46kb
27 28
sopA
sopB
30 31 32 33 34 35 36
antB
antA
icd
37
repA
38
cB
48 49
Circular
Linear

Figure 28-2 Map of N15 virion chromosome. The N15 linear virion chromosome is shown with a scale in kilo base pairs. Rectangles immediately above and below the scale represent predicted genes that, respectively, are transcribed rightward and leftward; their colors indicate similarity to known genes in the following way: genes that have been found in lambdoid phages (gray); genes that have been found in plasmids and non-lambdoid phages (black); no database match (white). The N15 gene names are given within or near the rectangles and alternate descriptive names are indicated above or below. Strongly predicted promoters (arrows in direction of transcription) and transcription terminators (T) are also indicated. Asterisks (*) mark the position of the centromere sites involved in plasmid partition, and small filled circles mark putative CB repressor binding sites. The lines at the bottom indicate the minimal sequences that are known to drive linear and circular plasmid replication in *E. coli*.

right arm genes and discussion of their possible functions can be found in (37).

Lysogeny Control

The extrachomosomal location of the N15 prophage apparently requires controlled expression of not only the repressor protein but also the genes responsible for prophage maintenance. In fact, analysis of N15 transcription patterns shows that about half the N15 genes are transcribed in the lysogen (37). This situation is radically different from that of λ lysogens (chapters 7 and 8) and suggests the possibility that phage N15 displays more complex regulatory mechanisms than phage λ. In order to identify genes involved in the control of lysogeny, lysogeny-defective mutants that form clear plaques were isolated. The mutations responsible for this inability to display lysogeny have been mapped to three distinct loci: *immA*, *immB*, and *immC* (44). Two of them — *immA* and *immB*, the phage N15 secondary and primary immunity regions, respectively — have been characterized.

Primary Immunity Region (*immB*)

Prophage superinfection immunity is encoded at *immB*, the primary immunity region, which was characterized by Lobocka et al. (21) and found to be structurally and functionally similar to the immunity region of lambdoid phages. *immB* contain three genes (figure 28-3): gene *38* (*cB*), gene *39* (*cro*), and gene *40* (*Q*). Gene *38* (*cB*) encodes a repressor protein which is similar and homologous to the phage λ *cI* gene. Clear-plaque mutants, mapping at *immB*, were found in the *cB* gene, supporting its role as a primary repressor. Gene *39* (*cro*) shows weak homology to *cro* genes of phages P22 and HK022 and occupies a position analogous to and similar in size to *cro*. The third gene, *40* (*Q*), encodes a protein that is similar to the transcription antitermination factor Q of phage φ82.

The *cB* gene is flanked by a complex array of divergent operator-promoter sites (figure 28-3). Lobocka et al. (21) identified operator sites and showed binding of the N15 CB protein to these sites in vitro, but positions of associated promoters are known exclusively from sequence analysis and require experimental verification. The two operators leftward of the *cB* gene overlap the predicted promoter of the N15 *repA* gene, implying that binding of the CB protein at these operators represses transcription of *repA*. This supposition is further supported by an observation that phage N15-based miniplasmids lacking the *cB* gene have a higher copy number than similar plasmids with an intact *cB* gene (29). The three operators rightward from *cB* overlap the predicted promoter of the *cB* gene as well as the predicted promoters of the "late" operon containing *cro* and *Q*. It was proposed (21) that the CB protein, by binding to these operators, represses both its own transcription and the transcription of *cro* and *Q*.

In addition to CB, two other factors were suggested to regulate the expression of *repA* (21), although these hypotheses have not been verified experimentally. The leader region of *repA* contains a sequence typical of strong rho-independent terminators, suggesting the involvement of termination and antitermination in the regulation of *repA*. Also in the leader region there is a putative counter-oriented promoter, P_{inc}, that is followed within approximately 80 bp by a strong terminator. This promoter could initiate transcription of a short RNA that is antisense to the leader sequence of *repA* and this short RNA may thereby modulate transcription of *repA*. Modulation of a replication gene's expression by antisense RNA is a common strategy employed by other plasmids and, particularly, by bacteriophage P1, where it regulates transcription of the lytic replicon (15) (chapter 24).

Secondary Immunity Region (*immA*)

Ravin et al. (32) have characterized the *immA* region (figure 28-4) and found that three open reading frames encode an inhibitor of cell division (coded by gene *icd*), an antirepressor protein (coded by gene *antA*), and a gene that may play an ancillary role in antirepression (*antB*). These genes may be transcribed from two promoters. The upstream promoter, *Pa*, could be repressed by the CB repressor, whereas the weaker downstream promoter, *Pb*, is constitutive. Full repression of the antirepressor operon is achieved by premature transcription termination at T1 and T2 that is elicited by a small RNA (CA RNA) produced by processing of the leader transcript of the operon — this mechanism is

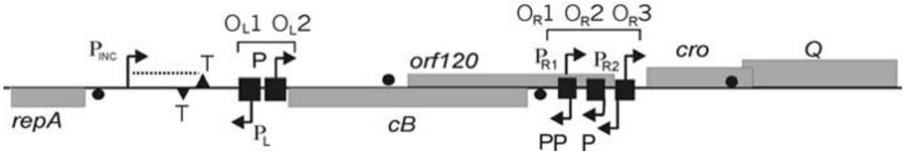


Figure 28-3 Organization of the *immB* region. Genes (gray rectangles), predicted promoters (P), terminators (T), operators (O), and ribosome-binding sites (filled circles) are indicated. Genes and sites shown above the main line apply to transcription from left to right (from right to left). The o*rf120* is preceded by promoter and putative ribosome-binding site but it has no homology to any known protein and its expression is uncertain.

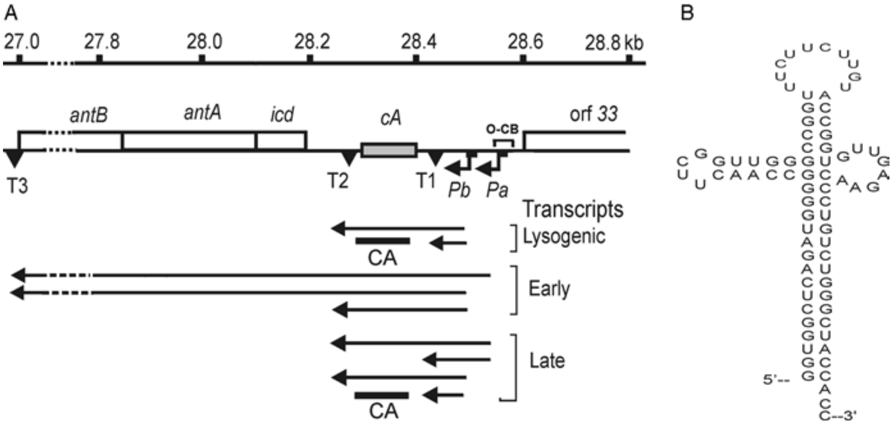


Figure 28-4 Organization of the *immA* region. A: Map and transcription map of the *immA* locus (32). Coordinates on the kb scale given according to the phage map. Protein-encoding genes are shown by open boxes, cA, which encodes an RNA, by gray box across the main line. Bent arrows indicate transcription start points corresponding to promoters *Pa* and *Pb* (solid boxes). Positions of putative terminators T1, T2, and T3 are shown by solid triangles. The transcripts of the *immA* region are indicated by arrows. The CA RNA is represented by a closed bar. B: Predicted secondary structure of N15 CA RNA.

similar to the one used in the anti-immunity system of phage P1 and the lysogeny control region of phage P4 (6, 7) (chapters 24 and 26). The CA RNA thus acts as a secondary repressor and clear-plaque mutants mapped at *immA* were found within the *cA* sequence. The CA RNA is a small, stable RNA molecule with a peculiar secondary structure: a double-stranded stalk, two stem-loops, and an 8 nucleotide single-stranded bulge (figure 28-4). The CA RNA appears to act as pseudo-antisense RNA since complementarity between specific sequences in CA RNA (specifically in the main loop and bulge) and corresponding sites on the untranslated leader transcript are required for efficient transcription termination.

The antirepressor functions encoded in the *immA* secondary immunity region determine the lysis–lysogeny decision of phage N15. Analysis of transcription patterns of the *immA* locus show that the structural genes (*icd*, *antA*, and *antB*) of the *ant* operon can only be expressed very soon after infection from the two promoters, before the CA RNA is produced by processing of the leader region of the transcript. The *ant* operon is then rapidly turned off once CA RNA is produced. In the lysogen, the CB repressor turns off the promoter *Pa*, while the second promoter, *Pb*, allows production of the CA RNA immunity factor. Synthesis of the CA RNA itself is negatively autoregulated (32) and the existence of two differently regulated promoters may be particularly important for phage N15 lytic development upon prophage induction. A transient inactivation of the CB repressor may allow transcription from the stronger promoter, *Pa*, and the existing CA RNA may not be sufficient to prevent transient transcription of the *ant* operon. This could result in the synthesis of antirepressor and a switch to the lytic pathway.

The first gene of the *ant* operon, *icd*, is not directly involved in antirepressor function. Expression of cloned *icd* instead leads to an immediate arrest of cell division, filamentation, and finally cell death (32). A possible role for this gene in the phage N15 life cycle could be to delay cell division soon after infection. On the one hand this may produce a "larger cell" that may be more advantageous for lytic development. On the other hand, in the case of the lysogenic pathway, it may provide sufficient time for expression of the N15 DNA replication and partitioning system so that each daughter cell will inherit the plasmid prophage once cell division resumes.

Segregational Sability of the Plasmid Pophage

The N15 plasmid prophage is maintained at three to five copies per bacterial chromosome and is very stable — its rate of spontaneous loss is less than 10^{-4} per generation (44). This is much less than would be expected in the case of random distribution of plasmid copies between bacterial daughter cells, and it implies the existence of special stabilization machinery. Two principal mechanisms ensuring stable inheritance of bacterial plasmids have been described: active partition of plasmid copies to daughter cells prior to division (for review see 18) and post-segregational killing of plasmid-free cells (reviewed in 12). It is unlikely that the first one is employed by N15 since prophage-free cells are easily accumulated at the nonpermissive temperature in N15 lysogens carrying "early" temperature-sensitive mutations (33).

Inspection of the complete nucleotide sequence of N15 revealed a region near the right end of the prophage with remarkable similarity to the *sop* locus, which governs partition of F plasmid copies to daughter cells. Ravin and Lane (28) demonstrated that the *sop* locus of N15 in fact determines stability of the prophage since Sop proteins, encoded at this locus, can stabilize the partition-defective N15 derivatives. The structural and functional organization of the N15 *sop* locus is similar to that of other partition loci including F *sop* and P1 *par.* These loci consist of a two-gene operon and an adjacent *cis*-acting site (see for review 18). The first gene (gene *28* = *sopA*) encodes a protein which binds to the promoter of the partition operon to repress transcription (the operon is thus negatively autoregulated) and which also acts directly in the partition process itself. The product of the second gene (gene *27* = *sopB*) binds to the *cis*-acting centromere site (C) to form a partition complex and acts as a corepressor of operon expression. The phage N15 and F-plasmid partition functions appear to be partly interchangeable: N15 SopA and SopB proteins can partly stabilize partition-defective mini-F plasmids and repress the F *sop* promoter, and vice versa. This work (28) revealed that the phage N15 partition system, although a functional analog of the F *sop* system, differs from it in several important respects:

1. The centromere site of phage N15 is *not* composed of a cluster of multiple inverted repeats adjacent to the *sop* operon, as is the case with F plasmids, the plasmid-forming phage P1, and other circular plasmids (except RK2). Instead, the centromere is represented by four inverted repeats located in different regions of the N15 genome. Each of these sites binds the SopB protein and acts as a centromere (14, 28).
2. Transcription of the F-plasmid *sop* operon is driven from one autoregulated promoter while transcription of phage N15 *sop* is driven by two major promoters (11). The first promoter is similar in sequence and function to the F sop promoter, and is repressed by Sop proteins. The second and stronger promoter is insensitive to regulation by Sop proteins but is tightly repressed by protelomerase, the N15 enzyme that completes prophage replication by generating hairpin telomeres. These promoters establish a regulatory link between the partition system and other processes of N15 maintenance.
3. The centromere sites are located in the regions of N15 genome that are supposed to be essential for replication and control of gene expression. One site, IR1, is located within the coding sequence of the replication gene, *repA*; the second, IR2, is located within gene *Q*, and the other two centromere sites, IR3 and IR4, are located close to the late promoters. This suggests that the N15 partition functions may be involved in the regulation of gene expression and replication, and this is further supported by an observation that the increased level of expression of the *sop* genes influences the copy number of an N15-based linear plasmid (N. Ravin and D. Lane, unpublished data). In two other systems the involvement of partition genes in the regulation of gene expression seems possible: ParB of phage P1 binds to the centromere site and is able to silence the nearby regions even if present at only physiological levels (40), and KorB of circular plasmid RK2 binds with different affinities to 12 sites within the RK2 plasmid (20).

These properties imply that the phage N15 *sop* system is not an independent functional unit but, like chromosomal partition functions, is part of a complex system involving replication and regulation of gene expression. Thus, study of the N15 partition system may provide key insights into the less well understood processes of chromosome partition.

Mechanism of Replication and Telomere Generation

The N15 Protelomerase

The N15 protelomerase was first hypothesized by Valentin Rybchin as an enzyme responsible for the formation of a linear-hairpin prophage molecule from the circularized phage DNA (figure 28-1). In this model the phage N15 protelomerase is a functional analog of lambdoid phage integrases. Sequencing of the N15 genome revealed an open reading frame whose translational product, gp29, has some sequence homology with the lambdoid phage integrases and the product of the BBB03 gene of *Borrelia burgdorferi.* This led to the suggestion that gene *29* (= *telN*) encodes the putative N15 protelomerase (see GenBank AF064539 annotation; 37; 41).

Deneke and coworkers cloned the N15 gene *29*, purified the corresponding protein, TelN, and demonstrated that it is responsible for processing the *telRL* site in vitro (9, 10). The enzyme cuts the target sequence, *telRL*, and joins the phosphodiester bonds, making hairpin ends in the absence of any further N15 or host-encoded proteins. The target 56 bp *telRL* site consists of the central 22 bp palindrome, *telO*, and two 14 bp flanking inverted repeats. The *telO* region has been predicted to provide B–Z DNA junctions, which might facilitate processing by TelN (41). DNase I footprinting of TelN–*telRL* complexes showed that an approximately 50 bp DNA segment is protected by TelN (10). Surface plasmon resonance studies demonstrated that two TelN molecules bind to *telRL*, suggesting that their concerted action is the basis for telomere resolution (10). Both linear and circular supercoiled DNA function as substrates for TelN, indicating that negative supercoiling is not required for the reaction. These results clearly show that the protelomerase can generate the linear prophage molecule after circularization of the infecting phage genome (figure28-1).

Ravin et al. (31) investigated the role of protelomerase in N15 prophage replication. They analyzed the protelomerase activity in vivo and demonstrated that this enzyme is required for replication of linear (but not circular) N15-based plasmids. Protelomerase was found to be an end-resolving enzyme responsible for processing of replicative intermediates. The authors demonstrated that the *telN* gene and *telRL* site constitute an independent functional unit acting on non-N15 replicons. Cloning of this module in circular mini-F and mini-P1 plasmids resulted in their linearization and further maintenance as linear plasmids with hairpin telomeres.

N15-based linear miniplasmids have been used as cloning vectors (29, 48) which, presumably due to the absence of supercoiling, appeared to be particularly suitable for cloning DNA sequences with inverted repeats (30). The functional independence of the protelomerase unit opens the prospect of constructing the linear derivatives of commonly used plasmid vectors.

Replication Mechanism

Little is known about the mechanism of replication of linear, covalently closed DNA in any biological system. It has been suggested (2) that the hairpin telomere is a potential solution to the problem that DNA polymerases alone cannot replicate the extreme ends of linear DNA molecules (50). Various models involving processing of replicative intermediates by an end-resolving enzyme have been proposed (figure 28-5; for a review see 3). These models could be discriminated into several principal alternatives: (i) location of the replication initiation site: internal versus telomere-proximal; (ii) direction of replication: uni- versus bi-directional, (iii) mode of replication: θ mechanism versus something else, and (iv) structure of a replicative intermediate processed by an end-resolving enzyme: circular dimer versus circular monomer versus linear molecule.

In order to identify the minimal set of genes able to drive replication of the N15 prophage, a set of miniplasmids consisting of different fragments of N15 DNA and an antibiotic resistance gene has been constructed (27). The shortest circular miniplasmid contained only gene *37* (*repA*), which is thus necessary and sufficient to drive replication of circular miniplasmid. The replication initiation site (*ori*) is located within the *repA* gene (27). The shortest constructed linear plasmid consists of *repA* and a protelomerase module (*telN* gene and *telRL* site). This is in agreement with the data that all replication mutations isolated so far have been mapped within *repA* (41, 42).

repA is a large gene, with a predicted protein product of 1324 amino acids. Sequence analysis of the putative RepA protein revealed motifs characteristic of bacterial and viral DNA primases and helicases, particularly of the phage P4 α replication protein (37) (for review of the phage P4 α protein, see chapter 26). These observations suggest that RepA may be a multi-domain protein similar to the replication proteins of bacterial plasmids replicating by the θ mechanism. Further arguing for the existence of primase and helicase activities of the RepA protein, phage N15 is able to productively infect *E. coli* strains carrying temperature-sensitive mutations in primase (*dnaG3*) and helicase

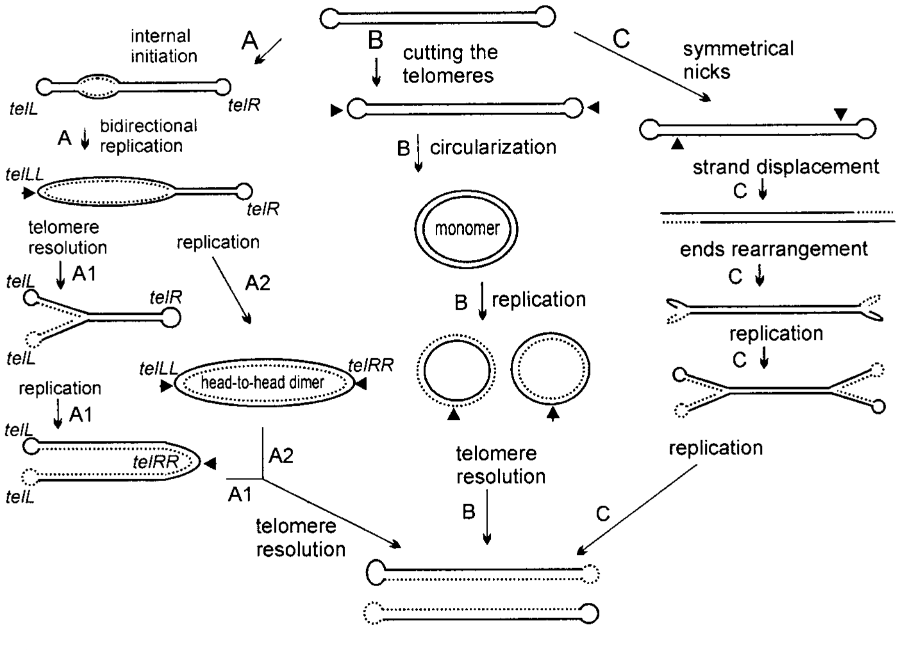


Figure 28-5 Model of N15 plasmid prophage replication (pathway A) and two other principal strategies to replicate linear DNAs with covalently closed ends (pathways B and C). A1 – protelomerase cuts before completion of replication, resulting in an Y-like structure. A2 – replication is finished before protelomerase cutting, circular head-to-head dimer. *telL*, *telR*, left and right telomere ends of N15 prophage; *telN*, protelomerase gene; *ori*, replication initiation site. Dotted lines indicate newly synthesized strands. Arrowheads indicate positions of cuts.

(*dnaB70*) genes at nonpermissive temperature (N. Ravin et al., manuscript in preparation). In addition, phage N15 replication was found to be independent of *E. coli polA*, *recA* (41), *dnaJ*, *dnaK*, and *grpE* (46) genes.

To determine the direction of plasmid replication we have integrated the phage N15 replication region, including *repA*, its promoter, and *cB* repressor gene, into the *E. coli* chromosome at the λ integration site. Direction of replication was determined by the analysis of which of the two nearby bacterial DNA markers flanking this site were amplified when the replicon was activated. We found (27) that both DNA markers are coordinately amplified, a result consistent with replication proceeding in a bidirectional fashion.

The next principal point allowing discrimination between different models of replication is the structure of the replicative intermediate processed by the end-resolving enzyme. Ravin and coworkers (31) constructed an N15 mutant carrying a deletion in the protelomerase gene and then cloned the *telN* gene into an expression vector under the control of a regulable promoter. The mutant may be maintained as a linear plasmid if the *telN* gene is expressed from the vector plasmid present in the same cell; repression of the *telN* gene results in accumulation of unprocessed replicative intermediates which were found to be circular head-to-head dimer molecules (31).

Electron microscopic analysis of intermediates generated in the course of replication of an N15-based linear plasmid allowed identification of three types of replicating molecules (27). Type 1 molecules, which are linear and the length of the linear plasmid, contain an internal “bubble” located near the position predicted for the *ori* site. These molecules likely represent an early step of internally initiated bidirectional replication. Type 2 molecules are circles located at the end of a linear DNA; these could result from replication to one end of the molecule without protelomerase resolution of the ends (most likely the left end, which is closer to the *ori* site than the right end). Type 3 molecules are Y-shaped molecules with two equal-length arms whose lengths are consistent with a single fork on the linear plasmid; these could result from TelN cleavage of the circle in a type 2 molecule. No circular molecules, either dimers or monomers, were found.

All these data suggest the following model of N15 plasmid replication (figure 28-5; reference 25), which is based on Bateman’s model of replication of palindromic telomeres (2), and is largely in agreement with the one proposed previously for the N15 prophage (41). Replication is initiated from an internal *ori* site, located within *repA*, follows the θ mode of DNA replication, and proceeds bidirectionally. After duplication of *telL*, protelomerase cuts this site creating hairpin ends, and thus a Y-shaped structure is formed. After the replication of the right telomere and subsequent cutting, two linear molecules are produced (figure 28-5, pathway A). Alternatively, under particular conditions, full replication of the molecule with the formation of a full head-to-head circular dimer may precede end resolution (figure 28-5, pathway B).

The above results suggest that N15 prophage replication could serve as a model for replication of other bacterial replicons with hairpin ends, such as the linear plasmids and chromosomes of *Borrelia burgdorferi*. Replication of these plasmids is initiated at an internal *ori* site and proceeds bidirectionally (24); sequence analysis has revealed regions of homology shared by the TelN protein and the *Borrelia* gene BBB03 product (37, 41). Chaconas et al. (5) showed that a synthetic sequence having the predicted structure of a replicated telomere functions as a substrate for telomere resolution in vivo and is sufficient to convert a circular replicon into a linear form. The authors suppose that the final step in the replication of *Borrelia* plasmids and chromosomes is a site-specific telomere breakage and reunion that occurred on the circular-dimer substrate. Later, Kobryn and Chaconas (19) showed that the BBB03 gene in fact encodes the telomere resolvase and demonstrated its activity in vitro. Thus, it seems likely that replication of *Borrelia* plasmids follows the same mode as that of phage N15. Genes highly similar to *telN* were also found in linear phage-plasmids PY54 (16) and phiKO2 (4). The cleavage-joining activities of PY54 protelomerase have also been shown in vitro (16).

Interestingly, the N15 model of replication differs from models suggested for eukaryotic replicons with hairpin ends. Particularly, poxvirus replication (8, 47) is initiated in the telomeric region, resulting in the formation of head-to-head and tail-to-tail concatemers through strand-displacement replication; the duplicated telomeres in concatemers are subsequently resolved by an as yet unknown enzyme to generate linear monomeric molecules with hairpin ends. These observations suggest the possibility of evolutionarily independent appearances of prokaryotic and eukaryotic replicons with covalently closed telomeres, rather than transfer of protelomerase genes from eukaryotes (poxviruses) to prokaryotes (*Borrelia*) or vice versa, as has been previously suggested (17).

Conclusions

Phage N15 is unique among bacteriophages in its genetic organization. Its morphogenetic genes and prophage repressor gene are similar to those of the lambdoid phage family. At the same time, N15 early operons contain mostly genes which reflect the prophage’s unique linear-plasmid lifestyle: the protelomerase gene, the antirepression operon, genes for replication and partition functions, and many others with as yet unknown functions. The specific mechanism of N15 prophage replication seems to be a combination of typical bidirectional θ-type strategy and action of the specific end-resolving enzyme, protelomerase. This mechanism could serve as a model for other replicons of such type. Phage N15 must have arisen either through a lambdoid progenitor's accumulation of new genetic modules from plasmid and bacterial sources or by an unknown plasmid that acquired a lambdoid set of "virion" genes. Therefore, N15 provides a very interesting model system for the study of phage and plasmid evolution as well as for the study of interactions between phages, plasmids and bacterial hosts.

Acknowledgments

I thank Victor Ravin and David Lane for critical reading of the manuscript. I acknowledge financial support from the US CRDF grant RB1–2043, INTAS grant 01–0786, RFBR grant 04-04-48643, and NATO grant LST.CLG.977113 "Functional domains in bacterial partition proteins".

References

1. Barbour, A. G., and C. F. Garon. 1987. Linear plasmids of the *Borrelia burgdorferi* have covalently closed ends. Science 237:409–411.
2. Bateman, A. 1975. Simplification of palindromic telomere theory. Nature 253:379.
3. Casjens, S. 1999. Evolution of the linear DNA replicons of the *Borrelia* spirochetes. Curr. Opin. Microbiol. 2:529–534.
4. Casjens, S. R., E. B. Gilcrease, W. M. Huang, K. L. Bunny, M. L. Pedulla, M. E. Ford, J. M. Hourtz, G. F. Hatfull, and R. W. Hendrix. 2004. The pKO2 linear plasmid prophage of *Klebsiella oxytoca.* J. Bacteriol. 186:1818–1832.
5. Chaconas, G., P. E. Stewart, K. Tilly, J. L. Bono, and P. Rosa. 2001. Telomere resolution in the Lyme disease spirochete. EMBO J. 20:3229–3327.
6. Citron, M., and H. Schuster. 1992. The c4 repressor of bacteriophage P1 is a processed 77 base antisense RNA. Nucleic Acids Res. 20:3085–3090.
7. Deho, G., S. Zangrossi, P. Sabbattini, G. Sironi, and D. Ghisotti. 1992. Bacteriophage P4 immunity controlled by small RNAs via transcription termination. Mol. Microbiol. 6:3415–3425.
8. DeMasi, J., S. Du, D. Lennon, and P. Traktman. 2001. Vaccinia virus telomeres: interaction with the viral I1, I6 and K4 proteins. J. Virol. 75:10090–10105.
9. Deneke, J., G. Ziegelin, R. Lurz, and E. Lanka. 2000. The protelomerase of temperate *Escherichia coli* phage N15 has cleaving-joining activity. Proc. Natl. Acad. Sci. USA 97:7721–7726.
10. Deneke, J., G. Ziegelin, R. Lurz, and E. Lanka. 2002. Phage N15 telomere resolution: target requirements for recognition and processing by the protelomerase. J. Biol. Chem. 277:10410–10419.
11. Dorokhov, B. D., D. Lane, and N. V. Ravin. 2003. Partition operon expression in the linear plasmid prophage N15 is controlled by both Sop proteins and protelomerase. Mol. Microbiol. 50:713–721.
12. Gerdes, K., J. S. Jacobsen, and T. Franch. 1997. Plasmid stabilization by post-segregational killing. Genet. Eng. (N.Y.) 19:49–61.
13. Golub, E. I., and V. K. Ravin. 1967. New system of phage mediated conversion. Dokl. Acad. Nauk. USSR 174:465–467.
14. Grigoriev, P .S., and M. B. Lobocka. 2001. Determinants of segregational stability of the linear plasmid-prophage N15 of *Escherichia coli.* Mol. Microbiol. 42: 355–368.
15. Heinrich, J., H. D. Riedel, B. Ruckert, R. Lurz, and H. Schuster. 1995. The lytic replicon of bacteriophage P1 is controlled by an antisense RNA. Nucleic Acids Res. 23:1468–1474.
16. Hertwig, S., L. Klein, R. Lurz, E. Lanka, and B. Appel. 2003. PY54, a linear plasmid prophage of *Yersinia enterocolitica* with covalently closed ends. Mol. Microbiol. 48:989–1003.
17. Hinnebusch, J., and K. Tilly. 1993. Linear plasmids and chromosomes in bacteria. Mol. Microbiol. 10:917–922.
18. Hiraga, S. 1992. Chromosome and plasmid partition in *Escherichia coli.* Annu. Rev. Biochem. 61:283–306.
19. Kobryn, K., and G. Chaconas. 2002. ResT, a telomere resolvase encoded by the Lyme disease spirochete. Mol. Cell 9:195–201.
20. Kostelidou, K., and C. M. Thomas. 2000. The hierarchy of KorB binding at its 12 binding sites on the broad-host-range plasmid RK2 and modulation of this binding by IncC1 protein. J. Mol. Biol. 295:411–422.
21. Lobocka, M., A. N. Svarchevsky, V. N. Rybchin, and M. Yarmolinsky. 1996. Characterization of the primary immunity region of the *Escherichia coli* linear plasmid prophage N15. J. Bacteriol. 178:2902–2910.
22. Malinin, A. Y., A. L. Vasilyev, G. A. Kholodiy, A. A. Vostrov, V. N. Rybchin, and A. N. Svarchevsky. 1992. Structure of the region in the genome of bacteriophage N15 necessary for hairpin formation on the ends of linear plasmid prophage. Mol. Gen. Mikrobiol. Virusol. 5–6:22–25.
23. Malinin, A. Y., A. A. Vostrov, V. N. Rybchin, and A. N. Svarchevsky. 1992. Structure of the linear plasmid N15 ends. Mol. Gen. Mikrobiol. Virusol. 5–6:19–22.
24. Picardeau, M., J. R. Lobry, and B. J. Hinnebusch. 1999. Physical mapping of an origin of bidirectional replication at the centre of the *Borrelia burgdorferi* linear chromosome. Mol. Microbiol. 32:437–445.
25. Ravin, N. V. 2003. Mechanisms of replication and telomere resolution of the linear plasmid prophage N15. FEMS Microbiol Lett. 221:1–6.
26. Ravin, N. V., O. I. Doroshenko, and V. K. Ravin. 1998. COS-region of the temperate phage N15. Mol. Gen. Mikrobiol. Virusol. 2:17–20.
27. Ravin, N. V., V. V. Kuprianov, E. B. Gilcrease, and S. R. Casjens. 2003. Bidirectional replication from an internal ori site of the linear N15 plasmid prophage. Nucleic Acids Res. 31:6552–6560.
28. Ravin, N., and D. Lane. 1999. Partition of the linear plasmid, N15: functional interactions with the *sop* locus of the F plasmid. J. Bacteriol. 181:6898–6906.
29. Ravin, N. V., and V. K. Ravin. 1994. An ultrahigh-copy plasmid based on the mini-replicon of the temperate phage N15. Mol. Gen. Mikrobiol. Virusol. 1:37–39.
30. Ravin, N. V., and V. K. Ravin. 1999. Use of a linear multicopy vector based on the mini-replicon of temperate coliphage N15 for cloning DNA with abnormal secondary structures. Nucleic Acids Res. 27: e13.

31. Ravin, N. V., T. S. Strakhova, and V. V. Kuprianov. 2001. The protelomerase of the phage-plasmid N15 is responsible for its maintenance in linear form. J. Mol. Biol. 312:899–906.
32. Ravin, N.V., A. N. Svarchevsky, and G. Dehò. 1999. The anti-imunity system of phage-plasmid N15: identification of the antirepressor gene and its control by a small processed RNA. Mol. Microbiol. 34:980–994.
33. Ravin, V. K. 1968. The functioning of the genes of temperate bacteriophage in lysogenic cells. Genetika 4:119–124.
34. Ravin, V. K. 1971. Lysogeny, p. 106. Nauka Press, Moscow.
35. Ravin, V. 1972. Lack of prophage N15 in *Escherichia coli* K12 chromosome. Genetika 8:189–191.
36. Ravin, V., and E. Golub. 1967. A study of phage conversion in *Escherichia coli*. I. The aquisition of resistance to bacteriophage T1 as a result of lysogenization. Genetika 4:113–121.
37. Ravin, V., N. Ravin, S. Casjens, M. Ford, G. Hatfull, and R. Hendrix. 2000. Genomic sequence and analysis of the atypical bacteriophage N15. J. Mol. Biol. 299:53–73.
38. Ravin, V. K., and M. G. Shulga. 1968. Temperature sensitive mutants of bacteriophage N15. Genetika 4:91–95.
39. Ravin, V. K., and M. G. Shulga. 1970. The evidence of extrachromosomal location of prophage N15. Virology 40:800–807.
40. Rodionov, O., M. Lobocka, and M. Yarmolinsky. 1999. Silencing of genes flanking the P1 plasmid centromere. Science 283:546–549.
41. Rybchin, V. N., and A. N. Svarchevsky. 1999. The plasmid prophage N15, a linear DNA with covalently closed ends. Mol. Microbiol. 33:895–903.
42. Sankova, T. P., A. N. Svarchevsky, and V. N. Rybchin. 1992. Isolation, characterization and mapping of N15 plasmid insertion mutants. Genetika 28:66–76.
43. Stoppel, R. D., M. Meyer, and H. G. Schlegel. 1995. The nickel resistance determinant cloned from the enterobacterium *Klebsiella oxytoca*: conjugational transfer, expression, regulation and DNA homologies to various nickel-resistant bacteria. Biometals 8:70–79.
44. Svarchevsky, A. N. and V. N. Rybchin. 1984. Characteristics of plasmid properties of bacteriophage N15. Mol. Gen. Mikrobiol. Virusol. 10:34–39.
45. Svarchevsky, A. N., and V. N. Rybchin. 1984. Physical mapping of plasmid N15 DNA. Mol. Gen. Mikrobiol. Virusol. 10:16–22.
46. Tilly, K. 1991. Independence of bacteriophage N15 lytic and linear plasmid replication from the heat shock proteins DnaJ, DnaK and GrpE. J. Bacteriol. 173: 6639–6642.
47. Traktman, P. 1996. Poxvirus DNA replication, pp. 775–798. *In* M. DePamphilis (ed.) DNA Replication in Eukaryotic Cells. Laboratory Press, Cold Spring Harbor, N.Y.
48. Vostrov, A. A., A. Y. Malinin, V. N. Rybchin, and A. N. Svarchevsky. 1992. Construction of linear plasmid vectors for cloning in *Escherichia coli* cells. Genetika 28:186–188.
49. Vostrov, A., O. Vostrukhina, A. Svarchevsky, and V. Rybchin. 1996. Proteins responsible for lysogenic conversion caused by coliphages N15 and phi80 are highly homologous. J. Bacteriol. 178:1484–1486.
50. Watson, J. 1972. Origin of concatemeric T7 DNA. Nature New Biol. 239:197–201.

29

Bacteriophage P22

PETER E. PREVELIGE, Jr.

Bacteriophage P22, a relative of bacteriophage λ, is a temperate phage of *Salmonella typhimurium* and has played key roles in the development of molecular biology. Originally isolated as a lysogen of *Salmonella*, it was with P22 that Zinder and Lederberg discovered the phenomenon of generalized transduction (113). In subsequent years, P22 has proven to be an important tool in *Salmonella* genetics, and equally importantly has served to illustrate key components of genetic and morphogenetic regulation (45). P22 shares many similarities in genetic structure and regulation with phage λ and there are excellent reviews of P22 biology (63, 88). The genome of P22 has been sequenced (11, 24, 61, 72, 100) and 65 genes have been annotated. The sequencing results support the hypothesis that phage P22 is a mosaic that has evolved through extensive recombination with other viruses. Significant progress has been made over the past 15 years in understanding the structural biology of P22 infection and replication which, after a brief overview, will be the focus of this chapter.

Life Cycle

Phage P22 enters the cell via the initial binding of the gp9 tailspike trimer to the O-antigen of the host lipopolysaccaride (2, 39, 40). Initial binding is followed by binding to a second receptor, and the subsequent ejection of the "E" or ejection proteins whose activity is required for active DNA to enter the cell (36, 38). Following infection, phage P22 can enter either a lytic or a lysogenic growth pathway. In the lytic pathway, viral replication proceeds immediately following infection and culminates approximately 1 hour later in the release of 300–500 phage progeny through cell lysis. In the lysogenic pathway, the phage chromosome integrates into the host chromosome and is passed to daughter cells during host cell division. The primary factor controlling this decision is the multiplicity of infection (moi). High moi favors lysogeny whereas low moi favors the lytic pathway (50). As will be discussed below, maintenance of the lysogenic state is an active process involving multiple repressor systems. Upon receipt of an appropriate trigger the prophage is excised and enters the lytic pathway.

Lytic Growth

In the case of the lytic cycle, the first genes to be expressed following DNA entry, are the "immediate early genes" which lie adjacent to the *c2* repressor gene and are transcribed by the host polymerase from the P_R, P_L and P_{ant} promoters (figure 29-1). These early genes code for functions involved in DNA replication, recombination, and regulation of gene expression. There are also two regulatory genes, genes *23* and *24*, both of which function as anti-terminators. Gene *24*, whose expression is driven off the P_L promoter acts as an anti-terminator similar in function to the λ phage *N* gene (35, 52, 62). In the absence of gene *24* function, the transcripts initiated from promoters P_L and P_R terminate before the early genes are transcribed. Given gene *24* expression, instead efficient transcription occurs. Similarly, the product of gene *23*, driven off the P_R promoter, is also an anti-terminator which anti-terminates transcription originating from the promoter, P_{LATE} (7, 52). The result of gene product 23 anti-termination is a 20,000 base transcript that encodes the genes required for phage assembly and release. The phage DNA itself is linear and is circularized by recombination and subsequently replicated by a rolling-circle mechanism. The resulting concatamer is packaged into the host by a headful mechanism which initiates at a specific site termed the "*pac*" site (14, 47, 74) (see chapter 6 for a general overview of DNA packaging by double-stranded DNA phage). Approximately 43,500 bp of DNA are packaged into the capsid (12). Since the genome comprises 41,724 bp in length, there is a terminal redundancy of approximately 4% in the packaged DNA. Packaging is processive with subsequent rounds beginning where the previous round terminated. The products of genes *2* and *3* are required for DNA packaging and act as a complex. Genetic evidence suggests that it is gp3 that initially recognizes the *pac* site (16, 107).

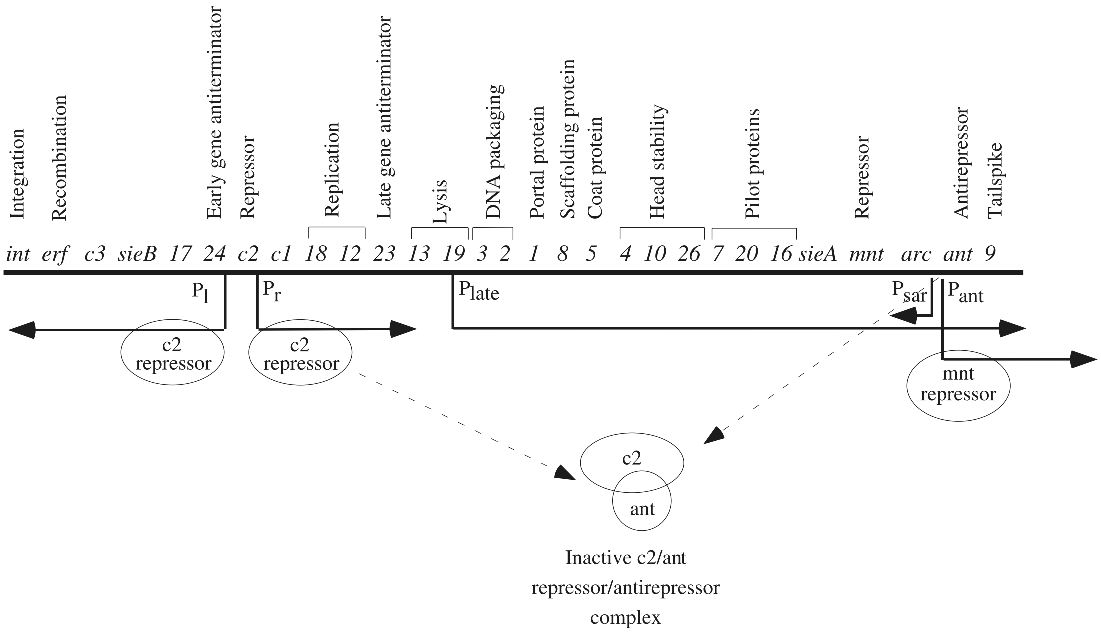

Integration
Recombination
Early gene antiterminator
Repressor
Replication
Late gene antiterminator
Lysis
DNA packaging
Portal protein
Scaffolding protein
Coat protein
Head stability
Pilot proteins
Repressor
Antirepressor
Tailspike
int erf c3 sieB 17 24 c2 c1 18 12 23 13 19 3 2 1 8 5 4 10 26 7 20 16 sieA mnt arc ant 9
P_l
P_r
P_{late}
P_{sar}
P_{ant}
c2 repressor
c2 repressor
mnt repressor
c2
ant
Inactive c2/ant repressor/antirepressor complex

Figure 29-1 The genetic map of bacteriophage P22. The genes are positioned above the line (not to scale), with their function indicated above. Promoter regions, and the loci of repressor action, are indicated below the line. The c2 repressor protein inhibits synthesis of the early genes, whose transcription is driven from the two promoters, P_L and P_R. Expression of the anti-terminators, 24 and 23, result in expression of the early and late genes, respectively. Notice that genes with related functions are clustered.

Generalized transduction is a result of packaging initiating at a *pac*-like site in the host cell chromosome with the resulting generation of a series of transducing particles (75, 76).

Lysogeny

In the lysogenic state, the P22 chromosome integrates into the host cell chromosome at a specific site, termed the *attB* site (49, 79). Integration is facilitated through the action of the protein product of the P22 *int* gene (77) with the assistance of a host-encoded integration host factor (IHF) whose function does not appear be essential (18). In the lysogenic state, relatively few P22 proteins are produced. The genes responsible for phage growth, and ultimately for host cell death, are repressed through the action of two repressor proteins, *c2* and *mnt*. The *c2* protein acts at the immunity region (*immC*) in a manner analogous to that of λ repressor (1, 20, 35). That is, it binds to O_L and O_R to prevent transcription of early genes. The unique *immI* region includes an antirepressor, *ant*, and three repressors: *mnt*, *arc*, and sar RNA. Mnt acts to repress the transcription of the Ant antirepressor (73, 101, 103). If *mnt* is turned off, Ant is synthesized. Ant binds to and inactivates the *c2* repressor protein with the result that the prophage enters the lytic pathway (8, 51, 87). The *ant* gene is transcribed from the strong P_{ant} promoter. The production of Ant itself is regulated through the action of the *arc* gene whose protein product binds within P_{ant}, thereby repressing Ant synthesis. In *arc* mutants high levels of Ant are produced, and these levels are sufficiently high to interfere with the synthesis of other essential phage proteins (86, 110).

The *ant* gene lies downstream of P_{LATE} and is additionally transcribed during the lytic cycle as part of the late operon. Despite transcription from the P_{LATE} promoter, the Ant protein is not subsequently synthesized. An additional promoter, P_{sar}, lies within the *ant* gene and directs the synthesis of a 69 nucleotide antisense RNA that binds the *ant* messenger RNA and inhibits translation (53, 108).

Both Mnt and Arc have been studied extensively as model DNA binding proteins. Mnt is functional as a tetramer and binds a 17 bp operator (101, 105). Mnt is comprised of two structural domains: a dimeric N-terminal domain, and a tetrameric C-terminal domain. The dimeric N-terminal domain is responsible for operator binding and specificity. Both *arc* and *mnt* are members of the ribbon–helix–helix family of transcription factors in which antiparallel β-sheet motifs, rather than helix–turn–helix motifs, are used to bind the operator DNA. In Mnt, the C-terminal tetramerization domain is comprised of two antiparallel α-helices with asymmetrical helical packing and a unique right-handed twist (57). Arc functions as a dimer and two dimers bind a 21 bp operator site in a highly cooperative reaction (9, 102). Detailed mutagenesis and folding studies on Arc have resulted, among other things, in a deeper understanding of the role of the hydrophobic core in protein stability (54), the importance of cooperative interactions in DNA binding (78), and the impact of effective concentration on protein stability (70).

Structure and Assembly

Significant advances in characterizing the structure and assembly of bacteriophage P22 have been made over the past 15 years. The assembly of infectious virions consists of two independent, linear pathways: the assembly of the capsid and the assembly of the tail. These two subassemblies then associate to form the infectious virion. Interest in the assembly of the capsid has stemmed from the observation that a "scaffolding" protein is required for proper form determination (15, 44). In the absence of scaffolding protein the coat protein subunits, which comprise the capsid, assemble into aberrant forms which are largely unclosed (22). The use of a scaffolding protein to guide assembly is not unique to bacteriophage P22. However, unlike the scaffolding protein of other phages, the scaffolding protein of phage P22 is not proteolytically cleaved during morphogenesis, a fact that significantly simplifies biochemical studies of assembly and maturation.

In the case of the tailspike protein, the properly folded and assembled form is SDS resistant whereas intermediates in folding and assembly are SDS-sensitive (28, 29). This fact has made possible quantitative investigations of protein folding in vivo using simple SDS-PAGE analysis as an analytical technique. Investigation of tailspike folding and scaffolding-protein-mediated P22 head assembly have provided mechanistic insight into protein folding and the control of biological self assembly.

Assembly of the Capsid

As is typical of double-stranded DNA phages, the P22 capsid assembles through the formation of a procapsid intermediate (43, 46). Approximately 300 molecules of the 33 kDa scaffolding protein co-assemble with 420 molecules of the 47 kDa coat protein to form a T = 7 procapsid (figure 29-2). During assembly one vertex is differentiated from the other 11 by the presence of a dodecameric complex of the 88 kDa portal protein (3, 4). This portal vertex serves as the conduit for DNA entry during packaging and egress during infection. In addition to the portal protein, approximately 12 copies each of three minor E or ejection proteins (gp7, gp16, and gp20) are incorporated into the procapsid. Although both their mechanism of action and location within the procapsid are unknown, these proteins are required for the delivery of an infectious viral genome to the host cell, and transit from the phage to the host cell during infection (36).

The 42 kb viral DNA is replicated as a concatemer by the rolling-circle mechanism and packaged in a headful fashion. Concomitant with DNA packaging the scaffolding protein exits the capsid and is capable of sequentially catalyzing approximately five rounds of assembly. DNA packaging induces a conformational change within the lattice resulting in head expansion and the appearance of angularity. The portal vertex is closed by the sequential addition of the products of genes *4*, *10*, and *26* and the stable capsid is now ready for tailspike addition (83).

Structure of the Capsid

The overall dimensions of the P22 procapsid and phage were determined by negative-stain electron microscopy as well as small-angle X-ray scattering by King and colleagues in the late 1970s (10, 21, 22). The P22 capsid lattice undergoes an expansion of approximately 10% from 580 nm to 610 nm

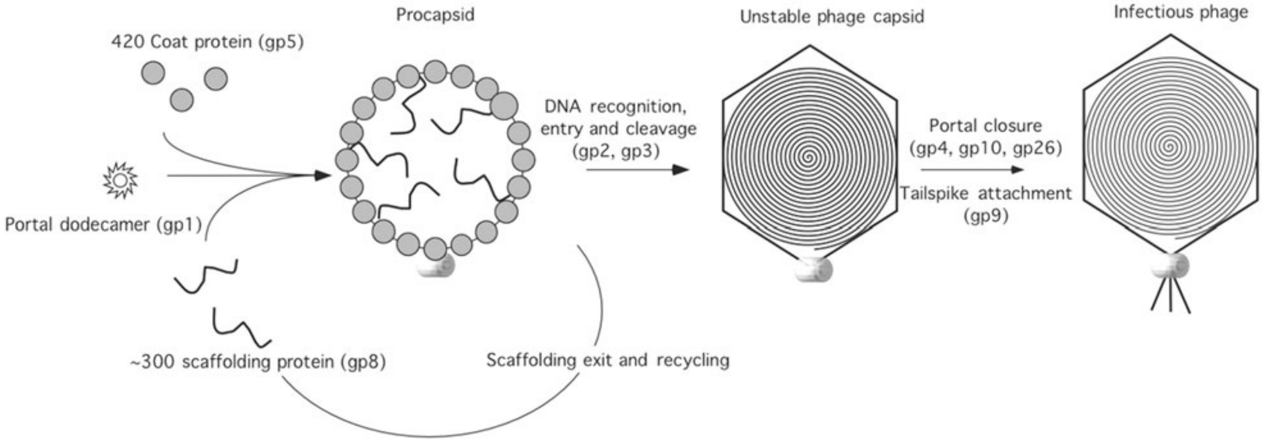

420 Coat protein (gp5)
Procapsid
Portal dodecamer (gp1)
~300 scaffolding protein (gp8)
Scaffolding exit and recycling
DNA recognition, entry and cleavage (gp2, gp3)
Unstable phage capsid
Portal closure (gp4, gp10, gp26)
Tailspike attachment (gp9)
Infectious phage

Figure 29-2 Morphogenetic pathway of P22. Approximately 420 molecules of the coat protein (gp5) co-assemble with approximately 300 molecules of the scaffolding protein (gp8). The minor (E) proteins, gp7, gp16, and gp20, as well as the portal protein (gp1), are incorporated during the early stages of assembly. The contameric double-stranded DNA is delivered to the portal vertex by the action of a complex composed of gp2 and gp3 and packaged in a headful mechanism. Scaffolding exit is concomitant with packaging. Once a headful of DNA is packaged (~104% of the genome), the DNA is cut and the portal vertex stabilized by the addtion of gp4, gp10, and gp26 (see chapter 6 for a general discussion of DNA packaging). Up to six tailspike trimers (gp9) are now able to attach, rendering the phage infectious.

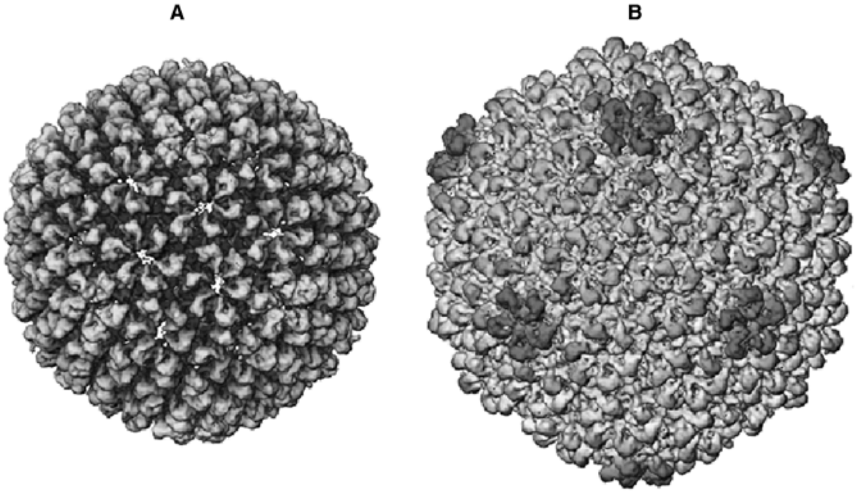


Figure 29-3 Cryo-electron micrographic image reconstructions. A: the P22 prohead. B: The mature virion. Both are hexameric and pentameric T = 7 lattices. The procapsid has a diameter of approximately 58 nm and is roughly spherical, while the phage has a diameter of 61 nm and displays pronounced angularity. Figure courtesy of Matthew Baker. See thebateriophages.org/frames_0290.htm for a color version of this figure.

during DNA packaging, as is typical for a double-stranded DNA phage. Both procapsid and capsid lattices appear to be T = 7 subtriangulated icosahedra based on freeze-fracture electron microscopy.

Subsequent years have seen the development of cryo-electron microscopy and image reconstruction as a powerful tool for the analysis of the structure of viral capsids. Based on cryo-electron microscopy image analysis, both procapsids and mature virions have been shown to be T = 7 lattices of hexamer- and pentamer-clustered coat-protein subunits as originally predicted by Caspar and Klug (17, 64). The procapsid form is indeed smaller and appears relatively spherical whereas the capsid lattice is larger and appears more angular (figure 29-3). A number of conformational changes within the lattice are responsible for the alterations in appearance. The procapsid lattice has holes, approximately 25 nm in diameter, located in the center of each hexameric coat-protein cluster. It has been postulated that these holes provide the exit port through which scaffolding leaves during DNA packaging (64). Furthermore, the coat-protein hexamers display a skewed character in the procapsid (93), which disappears leaving instead relatively symmetrical hexamers in the mature virion (112). This rearrangement appears to be accomplished through an outward movement of trimerically clustered subunits at the strict and local 3-fold axes (112), coupled with a rotational movement within the hexamer (42).

Cryo-electron microscopy provides the opportunity to visualize macromolecular disposition within the capsid as well as on the surface. Examination of procapsids revealed that the scaffolding was not globally icosahedrally ordered (93). However, comparison of procapsids and procapsid-like particles lacking scaffolding protein suggest that small regions of the scaffolding protein may contact the inner surface of the coat protein lattice at four of the six hexameric coat subunits (94). The observation that the scaffolding protein itself is not icosahedrally ordered within the procapsid, yet is responsible for modulating icosahedral assembly of the capsid, poses something of a dilemma. Where within the scaffolding protein does the information reside that determines icosahedral symmetry?

This problem is somewhat simplified by the observation that in the absence of scaffolding protein and at elevated temperature, the coat protein can polymerize into T = 4 as well as T = 7 procapsids. Cryo-electron microscopy reconstructions of these T = 4 particles indicate that the hexameric and pentameric clusters of coat protein are nearly identical in the two forms (95). This suggests that the role of the scaffolding protein is to modulate the relative positioning of hexamer- and pentamer-clustered coat protein into the appropriate lattice. However, as will be discussed below, in the case of phage P22 it does not appear that preformed hexamers and pentamers are the building blocks for assembly as has been suggested for phage HK97 (109).

Structure and Folding of the Coat Protein

While the procapsid, capsid, and coat protein subunits have proven refractory to crystallography to date, biochemical analysis has yielded insight into the structure of the coat protein. Circular dichroism and Raman spectroscopies indicate that the coat protein is a mixed α/β protein, composed of approximately 20% α-helix and 25% β-sheet (67, 90, 92). Proteolytic digestion studies indicate that the coat protein is composed of two structural domains of approximately equal size (48). The hinge region, which resides in the

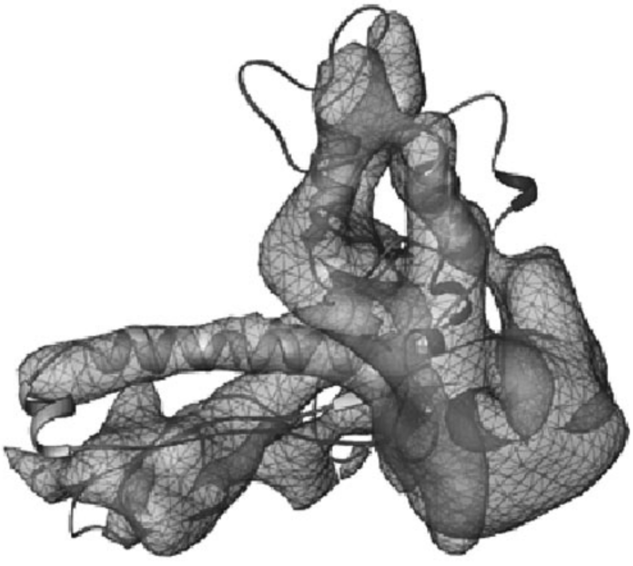


Figure 29-4 The cryo-electron micrographic structure of the coat protein. The cryo-electron microscopic electron density obtained from image reconstructions of P22 phage is shown in the cage structure and the fit of a ribbon diagram from the HK97 coat protein crystal structure is shown within the electron density map. See thebateriophages.org/frames_0290.htm for a color version of this figure.

residues 180–205, connects the two domains and becomes increasingly protected during capsid assembly and maturation (48, 96), a property shared with the capsid proteins of phages T4 and HK97 (41, 106). While there is no sequence homology, based on cryo-electron microscopy reconstructions it seems that the fundamental fold of these subunits as well as their packing within the capsid lattice is conserved between phages HK97 and P22 despite the fact that the phage P22 coat protein is substantially larger (429 residues versus 282) (figure 29-4) (42).

In complex macromolecular systems the boundaries between protein folding and assembly become blurred and consequently the P22 coat protein has been the focus of a number of folding studies. Temperature-sensitive mutations in coat proteins display a temperature-sensitive folding (tsf) phenotype in which they are temperature-labile during folding but display stability similar to wild-type once folded and assembled (30). At elevated temperatures, the *tsf* mutants form low-molecular-weight oligomers, (i.e., dimers, and trimers), which are incapable of assembly (89). The observation that 17 *tsf* mutations in the coat protein could be rescued by overexpression of the host chaperone, groEL, suggested chaperone-assisted folding (31). It proved possible to co-immunoprecipitate coat protein and groEL from cell lysates. In vivo experiments suggest that coat proteins partition between aggregation and assembly, and that the ratio of the fraction in each form can be altered by the presence or absence of groEL. Biophysical studies of the interaction of wild-type and mutant coat protein with groEL suggest that the subunits have considerable secondary and tertiary structure when bound to groEL. This in turn is suggestive of an interaction between groEL and the coat protein at a late folding stage. Interestingly, a cluster of amino acids near the C-terminus of the N-terminal domain act as intragenic global suppressor mutations that alter the balance between folding and aggregation.

The coat protein itself has unique biochemical properties. For example, even when properly folded the subunits are only marginally stable. Calorimetric studies indicate that that the melting temperature is near 37°C, a temperature where the phage itself is fully infectious. However, when assembled into the capsid lattice, the subunits are substantially stabilized and temperatures in excess of 80°C are required to denature both procapsid and phage lattices (26, 27). Thus, the subunits appear to be in a relatively flexible conformation, and derive the bulk of their stabilization from inter-lattice contacts. This property may confer the flexibility required both to adopt the "quasi-equivalent" conformations required to form a $T = 7$ lattice and to undergo the conformational changes required for expansion during DNA packaging. Hydrogen/deuterium exchange studies performed both with Raman and mass spectrometry have identified the regions of the coat protein which become stabilized upon assembly and maturation (96, 98). The expansion that accompanies packaging can be mimicked by gentle heating of the procapsids. While the transformation displays a high activation energy it interestingly is also exothermic, suggesting that the procapsid is a metastable, spring-loaded structure poised to rearrange to a lower energy form during DNA packaging (27).

Determining the structure of the packaged DNA within P22, or phage in general, has proven a formidable challenge and a number of models have been proposed. In the case of phage P22, Raman spectroscopic studies have indicated that the DNA remains in the B-form, and is uniformly curved rather than kinked (68, 92). Cryo-electron microscopy studies have demonstrated that the strands of the DNA are close-packed within the phage head with an average spacing of 2.5 nm. The coaxial spool model, as proposed by Earnshaw and Harrison (23), appears to be consistent with all extant data (112).

Structure of the Scaffolding Protein

The 33 kDa P22 scaffolding protein was predicted to be a highly α-helical protein, and this prediction has been borne out by experiment. Both circular dichroism spectroscopy and Raman spectroscopy suggest the presence of approximately 37% α-helix in the native protein (90, 92, 99). Hydrodynamic measurements suggest that the protein is a long rod-like molecule (25, 60). Both hydrogen/deuterium exchange studies, in which no exchange-protected core was observed, and calorimetric studies, in which no cooperative melting profile was observed, suggest that the molecule is highly flexible in solution (27, 97, 99). Analytical ultracentrifugation experiments demonstrated that the scaffolding protein is in a monomer–dimmer–tetramer equilibrium in solution. A naturally occurring mutation at residue 47 introduces a single cysteine into the protein. This mutant spontaneously forms dimers which are biologically active, suggesting that the dimeric form is likely to be the assembly-active species (60).

The C-terminal region of the scaffolding protein comprises at least part of the region that interacts with the coat protein. Deletion of the C-terminal 11 amino acids renders the scaffolding protein incapable of binding to coat protein, while scaffolding protein in which the N-terminus has been deleted up to residue 141 is capable of promoting procapsid assembly, albeit with somewhat reduced fidelity (58). The nulear magnetic resonance (NMR) structure of the C-terminal 35 residues of scaffolding protein revealed that the coat binding domain was a helix–loop–helix motif in which residues from each of the helical segments pack together to form a hydrophobic core (85). The fold of the coat binding domain of the scaffolding protein is structurally homologous to the tetratricopeptide repeat motif, a motif used to modulate protein–protein interactions. The surface of the coat binding domain is strongly positively charged, suggesting the likelihood of electrostatic binding. In accord with this suggestion, the presence of high salt concentrations blocks the binding of the scaffolding protein to preformed immature shells of coat protein (59).

While the C-terminal 35 residues (residues 269–303) are structured, the adjacent 31 residues (238–268) are highly flexible in solution. This result explains the cryo-electron miscroscopy data in which it was determined that the scaffolding protein is not icosahedrally disposed (93). Presumably the coat binding domain is anchored at the inner surface of the procapsid but the remainder of the scaffolding protein molecules are free to move around. It is not known whether the structure determined for the scaffolding protein represents the conformation when it is bound, or whether there is a conformational change in the scaffolding protein during the release step.

Scaffolding exit during DNA packaging appears to be an active process. Two mutations in the scaffolding protein, L177I and Q149W, result in failure of the scaffolding protein to be released and therefore these mutations block DNA packaging (32). This is a somewhat curious result because truncated versions of scaffolding protein, in which this region is missing entirely, are capable of both procapsid re-entry and exit in a model system.

Scaffolding Protein and Assembly

The development of an in vitro assembly system in which purified coat protein and scaffolding protein subunits could be induced to assemble into procapsid-like particles proved a significant advance in understanding the assembly of phage P22 and double-stranded DNA phage in general (66). The in vitro system recapitulates the need for scaffolding protein for the high-fidelity assembly seen in vivo. Importantly, for assembly this system requires only the simple mixing of coat and scaffolding protein without the need for solvent change such as pH or divalent cation alterations. Therefore, the components can be studied independently under assembly conditions.

As described above, the scaffolding protein exists in a monomer–dimmer–tetramer equilibrium in solution, a theme reiterated in other scaffolding proteins. The isolation of a mutant which spontaneously forms disulfide-crosslinked dimers (R74C) allowed for a direct demonstration of the activity of the dimeric form in vivo. Subsequent molecular biology-based dissection experiments indicated that the C-terminus of the scaffolding protein is required for coat protein binding and hence activity. The NMR-determined structure of the coat binding region revealed that this C-terminal region consists of a helix–turn–helix motif structurally analogous to the tetratricopeptide repeat motif (TPR) shown to be important for protein–protein interactions (figure 29-5). The helix–loop–helix is highly basic, and a significant component of the interaction between the coat and scaffolding proteins is electrostatic in nature.

One simple model by which scaffolding protein could control form determination is the "preformed core" model, in which scaffolding protein spontaneously associates to form a micellar-type structure whose outer surface is subsequently tiled by coat protein. Such a model has been proposed for T4 assembly based on the observation that cells infected with mutants that do not express coat protein accumulate scaffolding protein cores. However, this mechanism does not appear to be operative in P22. Under conditions where the scaffolding protein is active for assembly, no preformed cores can be detected in solution (60, 66).

A second model is one in which scaffolding protein is required only to initiate the assembly process. Subsequent growth steps can occur without the need for scaffolding protein. This model can be addressed by lowering the amount of scaffolding protein available for assembly, and determining the amount required for procapsid formation. This is a difficult experiment to perform in vivo because scaffolding protein appears to regulate its own level of synthesis at the post-transcriptional level, presumably by interacting with its own messenger RNA. However, this experiment can be performed in vitro, with the results indicating that a minimum of approximately 120 molecules of scaffolding protein are required for the assembly of a stable procapsid. Thus, it appears that scaffolding protein is required throughout the assembly process (66).

The kinetics of procapsid assembly have been studied in vitro to dissect the assembly pathway. Based on the presence of a lag phase, a critical concentration of coat

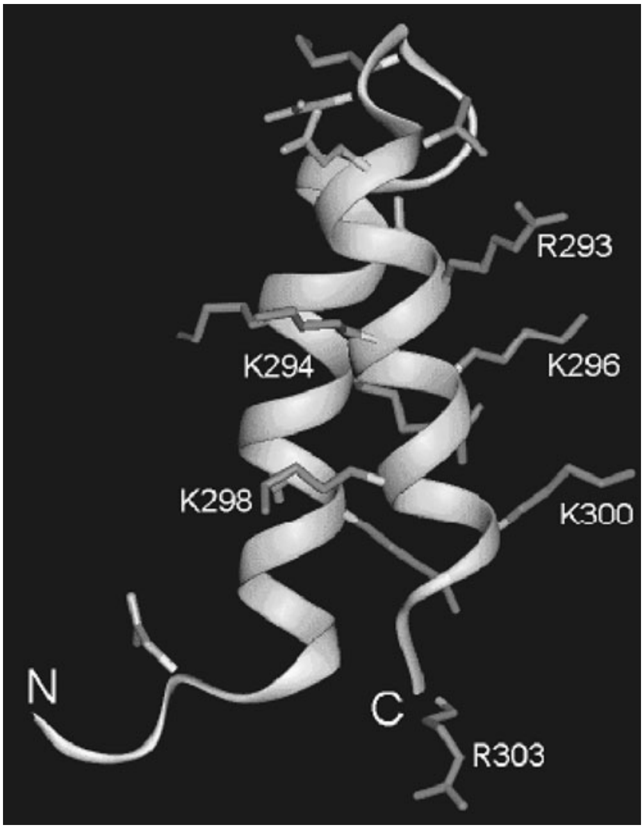


Figure 29-5 The solution structure of the C-terminus of the scaffolding protein as determined by NMR. The region displayed corresponds to residues 264–303. Note the two helices packed with a five-residue β-turn. Note also the basic lysine and arginine residues whose side chains point out into solution. See thebateriophages.org/frames_0290.htm for a color version of this figure.

protein for assembly, and the paucity of assembly intermediates, the overall assembly pathway appears nucleation-limited (65). A nucleus size of five coat protein subunits was deduced from the concentration dependence of the assembly reaction. Thus, it appears that nucleation of assembly occurs at a pentamer of coat protein, followed by outward growth. The growth appears to occur by the addition of coat protein monomers rather than by the addition of preassembled substructures such as pentamers, as has been proposed for phage HK97. It seems likely that this is a reflection of the relative strengths of the intersubunit interactions, and not a fundamental difference in the design of assembly pathways, which instead would seem likely to be conserved. It is not known whether subsequent subunits can add to every growing point, or if there is a preferred sequence of additions.

A biologically active procapsid contains not only coat and scaffolding protein but also a differentiated portal vertex and the ensemble of "E" proteins. The E protein gp16 can be incorporated into nascent procapsids in vitro if present early in the assembly reaction. This protein also is incorporated at the proper stoichiometric amounts even if supplied in excess. Together these observations suggest that only coat and scaffolding protein are required for controlled incorporation of the E protein. (91) Indeed, mutants in scaffolding protein have been isolated which do not incorporate gp16. In contrast, the gp1 portal protein has not been successfully incorporated into nascent procapsid in vitro whether supplied in monomeric or dodecameric form.

Based upon the kinetic analyses and structural studies, a model of how scaffolding protein might control form determination has been proposed. In this model, the role of the scaffolding protein is to stabilize otherwise unstable subunit additions by binding in key locations during assembly. The scaffolding protein thus may act to both stabilize and "energetically steer" the assembly process toward a T = 7 lattice (94).

Incorporation of the Portal

While there are a number of possible explanations for the failure of the portal protein to be incorporated during assembly in vitro, a likely one appears to be that an adaptor molecule is missing. The portal protein is a dodecameric structure that lies at a 5-fold vertex. Thus with P22, as with the other double-stranded DNA phage, there is a symmetry mismatch at the portal vertex. It has been proposed that the symmetry mismatch simplifies rotation because at no point during rotary motion would all the intersubunit interactions be either in phase or out of phase (34). A number of different adaptor molecules have been considered including messenger RNA and GroEL (6). At this point the existence of an adaptor molecule remains unproven.

The portal vertex presents a second level of asymmetry defined by the presence of a single portal vertex that is differentiated from the other 11 otherwise identical vertices. The mechanism by which P22 incorporates a single portal complex during assembly also remains obscure. In vivo virtually all proheads are capable of packaging DNA and converting to phage, implying the presence of a portal complex in nearly every procapsid-like particle. Virions with two portal complexes are extremely rare. The most straightforward way to insure the incorporation of one and only one portal complex is to couple portal incorporation to nucleation of assembly. However, mutant phage which do not express portal protein assemble procapsid-like particles at a rate identical to the assembly rate in the presence of portal protein (5). Thus, obligate incorporation during nucleation does not appear to be a viable model. A second model, in which the portal protein adds last, has been suggested for phage T7, and was explored by Moore and Prevelige (55). Pulse-chase experiments indicate that the portal protein cannot be added after prohead-like particles have been assembled. Thus, it appears that the portal protein is incorporated during the growth phase of assembly, perhaps with the help of an adaptor molecule.

Morphologically, the phage P22 portal vertex is similar to the portal protein complexes of other double-stranded DNA bacteriophage. It consists of a dodecamer of the portal subunits arranged in a ring with a diameter of approximately 180 nm. The ring has a central channel with a diameter of approximately 3 nm, which is of sufficient width to allow the passage of DNA (3). Spectroscopic studies reveal that it is a highly α-helical protein, and that it is possible to align predicted elements of secondary structure in the P22 portal protein with the core region of the known crystal structure of the phage Φ29 portal protein (56, 71). The P22 portal protein has been crystallized and crystallizes as a dodecamer (19). In contrast to Φ29, there is no evidence for RNA involvement with the P22 portal protein or DNA packaging.

Packaging of the concatemeric double-stranded DNA is modulated by the actions of the products of genes *2* and *3*, which together form a complex that is stabilized by ATP. Gp3, which corresponds to the terminase small subunit, is a 18.6 kDa protein and is responsible for recognition of the *pac* site (12–14, 16) The *pac* site itself is an asymmetrical 22 bp region that lies within gene *3*. Gp3 recognizes the *pac* site and makes a cleavage near it, generating a free end of DNA. This free end is then threaded into the procapsid through the portal complex until the head is full. At that point, the headful nuclease, presumably gp3, cleaves the DNA. Subsequent rounds of packaging begin at the newly generated end. The *pac* site corresponding to the first cleavage event does not lie at the end of the packaged DNA but rather approximately in the center of a 120 bp DNA end region (107). Whether this reflects the assembly of a complex of several molecule of gp2 and gp3, or slipping of the gp2/gp3 complex prior to cutting, is unclear. Gp2 is a 57.6 kDa protein which corresponds to the terminase large subunit and contains three predicted ATP binding sites. Gp2 is presumably involved in translocation of the DNA.

Structure of the Tailspike

Phage P22 infection of *Salmonella* begins with tailspike recognition of the O-antigen of the *Salmonella* lipopolysaccharide (LPS). The active tailspike is composed of a trimer of the 72 kDa product of gene *9* and three to six tailspike trimers bound to the phage capsid are required for infectivity (37, 38). The structures of a C-terminal fragment of tailspike (residues 109–666), an N-terminal fragment (residues 1–124), and the tailspike bound to the LPS have been solved crystallographically (80–82). As suggested by spectroscopic techniques, the tailspike is a highly β-sheet-rich protein and can be divided into four distinct domains (figure 29-6). The N-terminal 100 amino acids of each subunit are folded into an antiparallel β-sheet. Residues 143–540 form a 13-turn, right-handed, parallel β-coil. Resides 540–620 form an interdigitated β-sheet which then separates. The remaining amino acids are folded into individual C-terminal β-sheets. The overall appearance of the folded trimer is that of a fish, with the N-terminus forming the head, the parallel β-coil

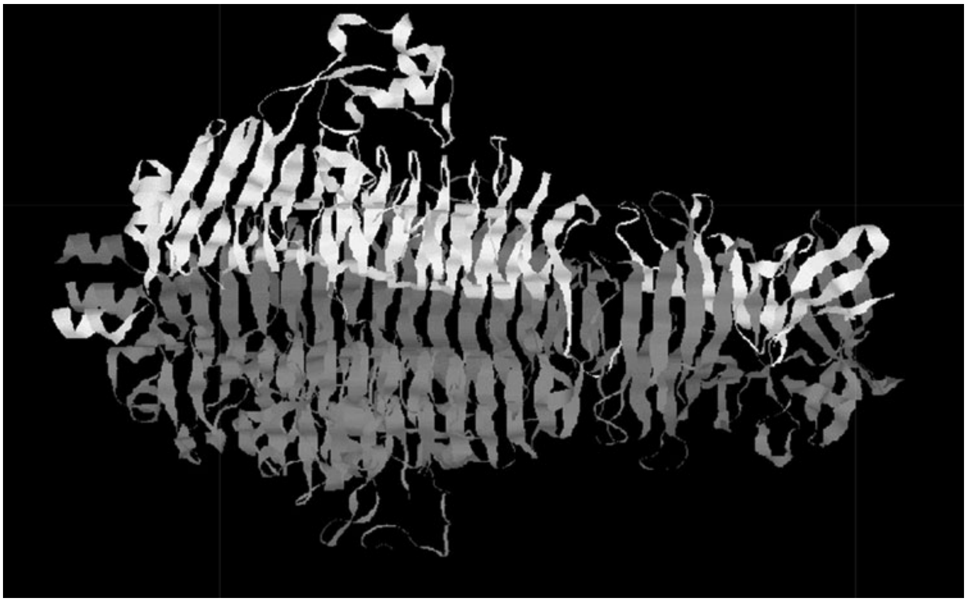


Figure 29-6 The crystal structure of the tailspike trimer. The structure of residues 109–666 (the C-terminus) of wild-type P22 tailspike protein is shown in ribbon-diagram form. The N-terminus of the molecule lies to the left of the figure, and the C-teminus lies to the right. The central region (the body of the fish) corresponds to the parallel β-coil structure. See thebateriophages.org/frames_0290.htm for an in-color version of this figure.

forming the body, and the C-terminal domain forming a caudal fin.

The binding site for the LPS is a cleft located in the central part of the β-coil which accommodates all eight carbohydrate residues of two repeating O-antigenic units (80). The tailspike protein displays receptor-destroying activity by cleaving the glycosidic bond between rhamnose and galactose. However, the enzymatic activity is slow, with the tailspike cleaving two bonds per minute at physiological temperature. Given the short time required for infection, it is unlikely that the enzymatic activity acts as a drill during infection. Rather, it is more likely that the enzymatic activity functions to facilitate the release of newly assembled phage from cell debris following cell lysis. A similar role has been proposed for the neuraminidase activity of influenza virus.

Folding of the Tailspike Protein

The unique biochemical properties of the tailspike protein, coupled with the well-defined genetics of the bacteriophage system, have made it an ideal subject for in vivo studies of protein folding. The tailspike folds in a multistep pathway, in which partially folded intermediates associate to form an SDS-sensitive protrimer. The protrimer then matures into a very stable SDS-insensitive trimer. The reason for this stability is that the central region of each subunit of the trimer is folded into a right-handed β-helix structure, and the C-terminal regions interdigitate to form β-sheet structures. The ability to readily identify and quantify the ratio of partially folded and fully folded protein by the simple technique of SDS-PAGE analysis has allowed for the isolation and characterization of a series of mutants which destabilize the partially folded intermediates during folding, the *tsf* mutants. Analysis of these mutants has lent support to the idea that protein folding and assembly takes advantage of transient interactions between amino acid residues that do not play critical roles in the stability of the final structure. At restrictive temperatures the *tsf* mutant proteins form aggregates, indicating that aggregation occurs through the interaction of partially folded protein molecules. These studies also indicate that alterations in the lifetime of the protein folding intermediates can result in aggregation (33, 84, 104, 111).

The tailspike protrimer represents a partially folded intermediate that can be isolated and studied biochemically. The tailspike protein contains eight cysteines per subunit, all of which are reduced in the mature folded tailspike trimer. Surprisingly, in the protrimer some of the cysteine residues are oxidized to form intermolecular disulfide bonds. The formation of these bonds in a folding intermediate may serve to facilitate subunit interactions or stabilize the protrimer during the reorganization required for successful folding (69).

The Future

The comparative study of the genetics of lamdoid phages has led to an appreciation of their evolutionary relatedness (chapter 27). It appears that this relatedness extends to the animal viruses as well. Comparative studies of the assembly pathways will serve to determine whether the pathways themselves are obligately conserved. For example, it may be that there are only certain strategies that will result in the formation of a biologically active, topologically closed viral capsid.

References

1. Ballivet, M., and H. Eisen. 1978. Purification and properties of phage P22 c2 repressor. Eur. J. Biochem. 82:175–180.
2. Baxa, U., S. Steinbacher, S. Miller, A. Weintraub, R. Huber, and R. Seckler. 1996. Interactions of phage P22 tails with their cellular receptor, Salmonella O-antigen polysaccharide. Biophys. J. 71:2040–2048.
3. Bazinet, C., J. Benbasat, J. King, J. M. Carazo, and J. L. Carrascosa. 1988. Purification and organization of the gene 1 portal protein required for phage P22 DNA packaging. Biochemistry 27:1849–1856.
4. Bazinet, C., and J. King. 1985. The DNA translocating vertex of double-stranded DNA bacteriophage. Annu. Rev. of Microbiol. 39:109–129.
5. Bazinet, C., and J. King. 1988. Initiation of P22 procapsid assembly in vivo. J. Mol. Biol. 202:77–86.
6. Bazinet, C., R. Villafane, and J. King. 1990. Novel second-site suppression of a cold-sensitive defect in phage P22 procapsid assembly. J. Mol. Biol. 216:701–716.
7. Botstein, D., C. H. Waddell, and J. King. 1973. Mechanism of head assembly and DNA encapsulation in *Salmonella* phage p22. I. Genes, proteins, structures and DNA maturation. J. Mol. Biol. 80:669–695.
8. Botstein, K., K. K. Lew, V. Jarvik, and C. A. Swanson. 1975. Role of antirepressor in the bipartite control of repression and immunity by bacteriophage P22. J. Mol. Biol. 91:439–462.
9. Bowie, J. U., and R. T. Sauer. 1989. Equilibrium dissociation and unfolding of the Arc repressor dimer. Biochemistry 28:7139–7143.
10. Casjens, S. 1979. Molecular organization of the bacteriophage P22 coat protein shell. J. Mol. Biol. 131:1–14.
11. Casjens, S., K. Eppler, R. Parr, and A. R. Poteete. 1989. Nucleotide sequence of the bacteriophage P22 gene 19 to 3 region: identification of a new gene required for lysis. Virology 171:588–598.
12. Casjens, S., and M. Hayden. 1988. Analysis in vivo of the bacteriophage P22 headful nuclease. J. Mol. Biol. 199:467–474.
13. Casjens, S., and W. M. Huang. 1982. Initiation of sequential packaging of bacteriophage P22 DNA. J. Mol. Biol. 157: 287–298.
14. Casjens, S., W. M. Huang, M. Hayden, and R. Parr. 1987. Initiation of bacteriophage P22 DNA packaging series.

Analysis of a mutant that alters the DNA target specificity of the packaging apparatus. J. Mol. Biol. 194:411–422.

15. Casjens, S., and J. King. 1974. P22 morphogenesis I. Catalytic scaffolding protein in capsid assembly. J. Supramol. Struct. 2:202–224.
16. Casjens, S., L. Sampson, S. Randall, K. Eppler, H. Wu, J. B. Petri, and H. Schmieger. 1992. Molecular genetic analysis of bacteriophage P22 gene 3 product, a protein involved in the initiation of headful DNA packaging. J. Mol. Biol. 227:1086–1099.
17. Caspar, D. L. D., and A. Klug. 1962. Physical principles in the construction of regular viruses. Cold Spring Harbor Symp. Quant. Biol. 27:1–24.
18. Cho, E. H., C. E. Nam, R. Alcaraz, Jr., and J. F. Gardner. 1999. Site-specific recombination of bacteriophage P22 does not require integration host factor. J. Bacteriol. 181:4245–4249.
19. Cingolani, G., S. D. Moore, P. E. Prevelige, Jr., and J. E. Johnson. 2002. Preliminary crystallographic analysis of the bacteriophage P22 portal protein. J. Struct. Biol. 139:46–54.
20. De Anda, J., A. R. Poteete, and R. T. Sauer. 1983. P22 c2 repressor. Domain structure and function. J. Biol. Chem. 258:10536–10542.
21. Earnshaw,W., S. Casjens, and S. C. Harrison. 1976. Assembly of the head of bacteriophage P22: X-ray diffraction from heads, proheads and related structures. J. Mol. Biol. 104:387–410.
22. Earnshaw, W., and J. King. 1978. Structure of phage P22 coat protein aggregates formed in the absence of the scaffolding protein. J. Mol. Biol. 126:721–747.
23. Earnshaw, W. C., and S. C. Harrison. 1977. DNA arrangement in isometric phage heads. Nature 268:598–602.
24. Eppler, K., E. Wyckoff, J. Goates, R. Parr, and S. Casjens. 1991. Nucleotide sequence of the bacteriophage P22 genes required for DNA packaging. Virology 183:519–538.
25. Fuller, M. T., and J. King. 1980. Regulation of coat protein polymerization by the scaffolding protein of bacteriophage P22. Biophys. J. 32:381–401.
26. Galisteo, M. L., C. L. Gordon, and J. King. 1995. Stability of wild-type and temperature-sensitive protein subunits of the phage P22 capsid. J. Biol. Chem. 270:16595–16601.
27. Galisteo, M. L., and J. King. 1993. Conformational transformations in the protein lattice of phage P22 procapsids. Biophys. J. 65:227–235.
28. Goldenberg, D., and J. King. 1982. Trimeric intermediate in the in vivo folding and subunit assembly of the tail spike endorhamnosidase of bacteriophage P22. Proc. Natl. Acad. Sci. USA 79:3403–3407.
29. Goldenberg, D. P., P. B. Berget, and J. King. 1982. Maturation of the tail spike endorhamnosidase of *Salmonella* phage P22. J. Biol. Chem. 257:7864–7871.
30. Gordon, C. L., and J. King. 1993. Temperature-sensitive mutations in the phage P22 coat protein which interfere with polypeptide chain folding. J. Biol. Chem. 268:9358–9368.
31. Gordon, C. L., S. K. Sather, S. Casjens, and J. King. 1994. Selective in vivo rescue by GroEL/ES of thermolabile folding intermediates to phage P22 structural proteins. J. Biol. Chem. 269:27941–27951.
32. Greene, B., and J. King. 1996. Scaffolding mutants identifying domains required for P22 procapsid assembly and maturation. Virology 225:82–96.
33. Haase-Pettingell, C., and J. King. 1997. Prevalence of temperature sensitive folding mutations in the parallel beta coil domain of the phage P22 tailspike endorhamnosidase. J. Mol. Biol. 267:88–102.
34. Hendrix, R. W. 1978. Symmetry mismatch and DNA packaging in large bacteriophages. Proc. Natl. Acad. Sci. USA 75:4779–4783.
35. Hilliker, S., and D. Botstein. 1975. An early regulatory gene of *Salmonella* phage P22 analogous to gene *N* of coliphage lambda. Virology 68:510–524.
36. Israel, V. 1977. E proteins of bacteriophage P22. I. Identification and ejection from wild-type and defective particles. J. Virol. 23:91–97.
37. Israel, V. 1978. A model for the adsorption of phage P22 to *Salmonella typhimurium*. J. Gen. Virol. 40:669–673.
38. Israel, V. 1976. Role of the bacteriophage P22 tail in the early stages of infection. J. Virol. 18:361–364.
39. Iwashita, S., and S. Kanegasaki. 1976. Enzymic and molecular properties of base-plate parts of bacteriophage P22. Eur. J. Biochem. 65:87–94.
40. Iwashita, S., and S. Kanegasaki. 1973. Smooth specific phage adsorption: endorhamnosidase activity of tail parts of P22. Biochem. Biophys. Res. Commun. 55:403–409.
41. Jardine, P. J., M. C. McCormick, C. Lutze-Wallace, and D. H. Coombs. 1998. The bacteriophage T4 DNA packaging apparatus targets the unexpanded prohead. J. Mol. Biol. 284:647–659.
42. Jiang, W., Z. Li, Z. Zhang, M. L. Baker, P. E. Prevelige, and W. Chiu. 2003. Coat protein fold and maturation transition of bacteriophage P22 seen at sub-nanometer resolution. Nat. Struct. Biol. 10:131–135.
43. King, J., D. Botstein, S. Casjens, W. Earnshaw, S. Harrison, and E. Lenk. 1976. Structure and assembly of the capsid of bacteriophage P22. Phil. Trans. R. Soc. Lond. B Biol. Sci. 276:37–49.
44. King, J., and S. Casjens. 1974. Catalytic head assembling protein in virus morphogenesis. Nature 251:112–119.
45. King, J., R. Griffin-Shea, and M. T. Fuller. 1980. Scaffolding proteins and the genetic control of virus shell assembly. Q. Rev. Biol. 55:369–393.
46. King, J., E. V. Lenk, and D. Botstein. 1973. Mechanism of head assembly and DNA encapsulation in *Salmonella* phage P22. II. Morphogenetic pathway. J. Mol. Biol. 80:697–731.
47. Kufer, B., H. Backhaus, and H. Schmieger. 1982. The packaging initiation site of phage P22. Analysis of packaging events by transduction. Mol. Gen. Genet. 187:510–515.
48. Lanman, J., R. Tuma, and P. E. Prevelige, Jr. 1999. Identification and characterization of the domain structure of bacteriophage P22 coat protein. Biochemistry 38: 14614–14623.
49. Leong, J. M., S. Nunes-Duby, C. F. Lesser, P. Youderian, M. M. Susskind, and A. Landy. 1985. The phi 80 and P22 attachment sites. Primary structure and interaction with *Escherichia coli* integration host factor. J. Biol. Chem. 260:4468–4477.

50. Levine, M. 1957. Mutations in the temperate phage P22 and lysogeny in *Salmonella*. Virology 3:22–41.
51. Levine, M., S. Truesdell, T. Ramakrishnan, and M. J. Bronson. 1975. Dual control of lysogeny by bacteriophage P22: an antirepressor locus and its controlling elements. J. Mol. Biol. 91:421–438.
52. Lew, K., and S. Casjens. 1975. Identification of early proteins coded by bacteriophage P22. Virology 68:525–533.
53. Liao, S. M., T. H. Wu, C. H. Chiang, M. M. Susskind, and W. R. McClure. 1987. Control of gene expression in bacteriophage P22 by a small antisense RNA. I. Characterization in vitro of the Psar promoter and the sar RNA transcript. Genes Dev. 1:197–203.
54. Milla, M. E., B. M. Brown, and R. T. Sauer. 1994. Protein stability effects of a complete set of alanine substitutions in Arc repressor. Nat. Struct. Biol. 1:518–523.
55. Moore, S. D., and P. E. Prevelige, Jr. 2002. Bacteriophage P22 portal vertex formation in vivo. J. Mol. Biol. 315:975–994.
56. Moore, S. D., and P. E. Prevelige, Jr. 2001. Structural transformations accompanying the assembly of bacteriophage P22 portal protein rings in vitro. J. Biol. Chem. 276:6779–6788.
57. Nooren, I. M., R. Kaptein, R. T. Sauer, and R. Boelens. 1999. The tetramerization domain of the Mnt repressor consists of two right-handed coiled coils. Nat. Struct. Biol. 6:755–759.
58. Parker, M. H., S. Casjens, and P. E. Prevelige, Jr. 1998. Functional domains of bacteriophage P22 scaffolding protein. J. Mol. Biol. 281:69–79.
59. Parker, M. H., and P. E. Prevelige, Jr. 1998. Electrostatic interactions drive scaffolding/coat protein binding and procapsid maturation in bacteriophage P22. Virology 250:337–349.
60. Parker, M. H., W. F. Stafford, 3rd, and P. E. Prevelige, Jr. 1997. Bacteriophage P22 scaffolding protein forms oligomers in solution. J. Mol. Biol. 268:655–665.
61. Pedulla, M. L., M. E. Ford, T. Karthikeyan, J. M. Houtz, R. Hendrix, G. F. Hatfull, A. R. Poteete, E. B. Gilcrease, D. A. Winn-Stapley, and S. Casjens. 2003. Corrected sequence of the bacteriophage P22 genome. J. Bacteriol. 185:1475–1477.
62. Pipas, J. M., and R. H. Reeves. 1977. Patterns of transcription in bacteriophage P22-infected *Salmonella typhimurium*. J. Virol. 21:825–828.
63. Poteete, A. R. 1988. Bacteriophage P22, pp. 647–682. *In* R. Calendar (ed.) The Bacteriophages, vol. II. Plenum Press, New York.
64. Prasad, B. V., P. E. Prevelige, E. Marietta, R. O. Chen, D. Thomas, J. King, and W. Chiu. 1993. Three-dimensional transformation of capsids associated with genome packaging in a bacterial virus. J. Mol. Biol. 231:65–74.
65. Prevelige, P. E., Jr., D. Thomas, and J. King. 1993. Nucleation and growth phases in the polymerization of coat and scaffolding subunits into icosahedral procapsid shells. Biophys. J. 64:824–835.
66. Prevelige, P. E., Jr., D. Thomas, and J. King. 1988. Scaffolding protein regulates the polymerization of P22 coat subunits into icosahedral shells in vitro. J. Mol. Biol. 202:743–757.
67. Prevelige, P. E., Jr., D. Thomas, J. King, S. A. Towse, and G. J. Thomas, Jr. 1990. Conformational states of the bacteriophage P22 capsid subunit in relation to self-assembly. Biochemistry 29:5626–5633.
68. Reilly, K. E., and G. J. Thomas, Jr. 1994. Hydrogen exchange dynamics of the P22 virion determined by time-resolved Raman spectroscopy. Effects of chromosome packaging on the kinetics of nucleotide exchanges. J. Mol. Biol. 241:68–82.
69. Robinson, A. S., and J. King. 1997. Disulphide-bonded intermediate on the folding and assembly pathway of a non-disulphide bonded protein [see comments]. Nat. Struct. Biol. 4:450–455.
70. Robinson, C. R., and R. T. Sauer. 1996. Equilibrium stability and sub-millisecond refolding of a designed single-chain Arc repressor. Biochemistry 35:13878–13884.
71. Rodriguez-Casado, A., S. D. Moore, P. E. Prevelige, Jr., and G. J. Thomas, Jr. 2001. Structure of bacteriophage P22 portal protein in relation to assembly: investigation by Raman spectroscopy. Biochemistry 40:13583–13591.
72. Sampson, L., and S. Casjens. 1993. Nucleotide sequence of *Salmonella* bacteriophage P22 head completion genes 10 and 26. Nucleic Acids Res. 21:3326.
73. Sauer, R. T., W. Krovatin, J. DeAnda, P. Youderian, and M. M. Susskind. 1983. Primary structure of the immI immunity region of bacteriophage P22. J. Mol. Biol. 168:699–713.
74. Schmieger, H. 1984. *pac* sites are indispensable for in vivo packaging of DNA by phage P22. Mol. Gen. Genet. 195:252–255.
75. Schmieger, H. 1972. Phage P22-mutants with increased or decreased transduction abilities. Mol. Gen. Genet. 119:75–88.
76. Schmieger, H., and H. Backhaus. 1976. Altered cotransduction frequencies exhibited by HT-mutants of *Salmonella*-phage P22. Mol. Gen. Genet. 143:307–309.
77. Smith, H. O., and M. Levine. 1967. A phage P22 gene controlling integration of prophage. Virology 31:207–216.
78. Smith, T. L., and R. T. Sauer. 1995. P22 Arc repressor: role of cooperativity in repression and binding to operators with altered half-site spacing. J. Mol. Biol. 249:729–742.
79. Smith-Mungo, L., I. T. Chan, and A. Landy. 1994. Structure of the P22 *att* site. Conservation and divergence in the lambda motif of recombinogenic complexes. J. Biol. Chem. 269:20798–20805.
80. Steinbacher, S., U. Baxa, S. Miller, A. Weintraub, R. Seckler, and R. Huber. 1996. Crystal structure of phage P22 tailspike protein complexed with *Salmonella* sp. O-antigen receptors. Proc. Natl. Acad. Sci. USA 93:10584–10588.
81. Steinbacher, S., S. Miller, U. Baxa, N. Budisa, A. Weintraub, R. Seckler, and R. Huber. 1997. Phage P22 tailspike protein: crystal structure of the head-binding domain at 2.3 Å, fully refined structure of the endorhamnosidase at 1.56 Å resolution, and the molecular basis of O-antigen recognition and cleavage. J. Mol. Biol. 267:865–880.
82. Steinbacher, S., R. Seckler, S. Miller, B. Steipe, R. Huber, and P. Reinemer. 1994. Crystal structure of P22 tailspike protein: interdigitated subunits in a thermostable trimer. Science 265:383–386.
83. Strauss, H., and J. King. 1984. Steps in the stabilization of newly packaged DNA during phage P22 morphogenesis. J. Mol. Biol. 172:523–543.

84. Sturtevant, J. M., M. H. Yu, C. Haase-Pettingell, and J. King. 1989. Thermostability of temperature-sensitive folding mutants of the P22 tailspike protein. J. Biol. Chem. 264:10693–10698.
85. Sun, Y., M. H. Parker, P. Weigele, S. Casjens, P. E. Prevelige, Jr., and N. R. Krishna. 2000. Structure of the coat protein-binding domain of the scaffolding protein from a double-stranded DNA virus. J. Mol. Biol. 297:1195–1202.
86. Susskind, M. M. 1980. A new gene of bacteriophage P22 which regulates synthesis of antirepressor. J. Mol. Biol. 138:685–713.
87. Susskind, M. M., and D. Botstein. 1975. Mechanism of action of *Salmonella* phage P22 antirepressor. J. Mol. Biol. 98:413–424.
88. Susskind, M. M., and D. Botstein. 1978. Molecular genetics of bacteriophage P22. Microbiol. Rev. 42:385–413.
89. Teschke, C. M., and J. King. 1995. In vitro folding of phage P22 coat protein with amino acid substitutions that confer in vivo temperature sensitivity. Biochemistry 34:6815–6826.
90. Teschke, C. M., J. King, and P. E. Prevelige, Jr. 1993. Inhibition of viral capsid assembly by 1,1′-bi(4-anilinonaphthalene-5-sulfonic acid). Biochemistry 32:10658–10665.
91. Thomas, D., and P. Prevelige, Jr. 1991. A pilot protein participates in the initiation of P22 procapsid assembly. Virology 182:673–681.
92. Thomas, G. J., Jr., Y. Li, M. T. Fuller, and J. King. 1982. Structural studies of P22 phage, precursor particles, and proteins by laser Raman spectroscopy. Biochemistry 21:3866–3878.
93. Thuman-Commike, P. A., B. Greene, J. Jakana, B.V. Prasad, J. King, P. E. Prevelige, Jr., and W. Chiu. 1996. Three-dimensional structure of scaffolding-containing phage p22 procapsids by electron cryo-microscopy. J. Mol. Biol. 260:85–98.
94. Thuman-Commike, P. A., B. Greene, J. A. Malinski, M. Burbea, A. McGough, W. Chiu, and P. E. Prevelige, Jr. 1999. Mechanism of scaffolding-directed virus assembly suggested by comparison of scaffolding-containing and scaffolding-lacking P22 procapsids. Biophys. J. 76:3267–3277.
95. Thuman-Commike, P. A., B. Greene, J. A. Malinski, J. King, and W. Chiu. 1998. Role of the scaffolding protein in P22 procapsid size determination suggested by T = 4 and T = 7 procapsid structures. Biophys. J. 74:559–568.
96. Tuma, R., L. U. Coward, M. C. Kirk, S. Barnes, and P. E. Prevelige, Jr. 2001. Hydrogen–deuterium exchange as a probe of folding and assembly in viral capsids. J. Mol. Biol. 306:389–396.
97. Tuma, R., M. H. Parker, P. Weigele, L. Sampson, Y. Sun, N. R. Krishna, S. Casjens, G. J. Thomas, Jr., and P. E. Prevelige, Jr. 1998. A helical coat protein recognition domain of the bacteriophage P22 scaffolding protein. J. Mol. Biol. 281:81–94.
98. Tuma, R., P. E. Prevelige, Jr., and G. J. Thomas, Jr. 1998. Mechanism of capsid maturation in a double-stranded DNA virus. Proc. Natl. Acad. Sci. USA 95:9885–9890.
99. Tuma, R., P. E. Prevelige, Jr., and G. J. Thomas, Jr. 1996. Structural transitions in the scaffolding and coat proteins of P22 virus during assembly and disassembly. Biochemistry 35:4619–4627.
100. Vander Byl, C., and A. M. Kropinski. 2000. Sequence of the genome of *Salmonella* bacteriophage P22. J. Bacteriol. 182:6472–6481.
101. Vershon, A. K., S. M. Liao, W. R. McClure, and R. T. Sauer. 1987. Bacteriophage P22 Mnt repressor. DNA binding and effects on transcription in vitro. J. Mol. Biol. 195:311–322.
102. Vershon, A. K., S. M. Liao, W. R. McClure, and R. T. Sauer. 1987. Interaction of the bacteriophage P22 Arc repressor with operator DNA. J. Mol. Biol. 195:323–331.
103. Vershon, A. K., P. Youderian, M. M. Susskind, and R. T. Sauer. 1985. The bacteriophage P22 arc and mnt repressors. Overproduction, purification, and properties. J. Biol. Chem. 260:12124–12129.
104. Villafane, R., and J. King. 1988. Nature and distribution of sites of temperature-sensitive folding mutations in the gene for the P22 tailspike polypeptide chain. J. Mol. Biol. 204:607–619.
105. Waldburger, C. D., and R. T. Sauer. 1995. Domains of Mnt repressor: roles in tetramer formation, protein stability, and operator DNA binding. Biochemistry 34:13109–13116.
106. Wikoff, W. R., R. L. Duda, R. W. Hendrix, and J. E. Johnson. 1999. Crystallographic analysis of the double-stranded DNA bacteriophage HK97 mature empty capsid. Acta. Crystallogr. D Biol. Crystallogr. 55:763–771.
107. Wu, H., L. Sampson, R. Parr, and S. Casjens. 2002. The DNA site utilized by bacteriophage P22 for initiation of DNA packaging. Mol. Microbiol. 45:1631–1646.
108. Wu, T. H., S. M. Liao, W. R. McClure, and M. M. Susskind. 1987. Control of gene expression in bacteriophage P22 by a small antisense RNA. II. Characterization of mutants defective in repression. Genes Dev. 1:204–212.
109. Xie, Z., and R. W. Hendrix. 1995. Assembly in vitro of bacteriophage HK97 proheads. J. Mol. Biol. 253:74–85.
110. Youderian, P., S. J. Chadwick, and M. M. Susskind. 1982. Autogenous regulation by the bacteriophage P22 arc gene product. J. Mol. Biol. 154:449–464.
111. Yu, M. H., and J. King. 1988. Surface amino acids as sites of temperature-sensitive folding mutations in the P22 tailspike protein. J. Biol. Chem. 263:1424–1431.
112. Zhang, Z., B. Greene, P. A. Thuman-Commike, J. Jakana, P. E. Prevelige, Jr., J. King, and W. Chiu. 2000. Visualization of the maturation transition in bacteriophage P22 by electron cryomicroscopy. J. Mol. Biol. 297:615–626.
113. Zinder, N. D., and J. Lederberg. 1952. Genetic exchange in *Salmonella*. J. Bacteriol. 64:679–699.

30

The Bacteriophage Mu

LUCIANO PAOLOZZI
PATRIZIA GHELARDINI

Mu was the first mobile genetic element identified in prokaryotes. Since its first isolation, in 1963, it has attracted the interest of many biologists. This interest is a consequence of its double nature: Mu is both a bacteriophage and a transposon.

Studies on Mu began when L. Taylor published a paper on a new temperate bacteriophage that showed the unusual ability to insert its DNA in multiple sites of the host genome. This integration was coupled with the induction of mutations in the host cell (236). The insertional mutagenesis phenomenon was thus described for the first time and it was shown that when sensitive bacteria (*Escherichia coli* K12) were infected with bacteriophage Mu, they exhibited increased mutation rates at many genetic loci. The prophage was always linked to the chromosomal site of the induced mutation. These properties were similar to those of the "controlling elements" postulated by Barbara McClintock (176), consisting of DNA moving from one position to another in the maize chromosome that resulted in modification and even suppression of the function of some genes.

Soon after these interesting discoveries, phage Mu revealed other fascinating aspects: (i) an unusual DNA structure consisting of heterogeneous host sequences at both ends of the Mu chromosome (27, 48, 49, 110), (ii) an ability to alternate host range by inversion of a DNA segment (127, 247), (iii) a paradigmatic mode of DNA replication by transposition, (iv) formation of a variety of host DNA rearrangements in the bacterial chromosome including inversions, duplications, and deletions of adjacent genes, replicon fusions, and transpositions of host DNA segments (57, 59, 60, 61, 160, 220, 241, 243), and (v) regulation, by DNA methylation, of the expression of the modification gene, *mom* (95, 96, 102, 119, 123, 205). As a transposon, Mu shows another unusual peculiarity: its ability to exploit two transposition mechanisms, replicative and nonreplicative, both of which are mediated by the same recombinase, the Mu transposase (the product of Mu gene *A*) (31). The Mu ability to promote genome rearrangements during its lytic cycle, in turn, was largely exploited in genetic manipulations for the construction of a large collection of bacterial strains, particularly in Ariane Toussaint and Malcolm Casadaban's laboratories. As befitting an organism with such an important and varied life-style, there have been a number of previous reviews of Mu (153, 185), including a chapter in the previous edition of this book, published in 1988 (92), and even an entire book devoted to phage Mu, which was published in 1987 (235). Here we strive to bring the reader up to date on our understanding of biology of this fascinating virus.

Overview of Mu

The Mu Life Cycle

The life cycle of Mu starts with adsorption and injection of its linear, double-stranded DNA genome into a sensitive host (figure 30-1). The injected DNA is flanked by a small region whose length and sequence vary in each viral particle and which constitutes a vestige of the host chromosome where the phage previously developed. Together with the DNA, a phage-encoded protein is also injected, the Mu N protein (product of gene *N*). This protein is necessary to convert the linear form of the phage genome into a circular, noncovalently closed form. Subsequently, in a process known as nonreplicative transposition (which is mediated through the action of the transposase, the Mu A protein) the two strands of phage DNA, but not the bacterial sequences associated with them, are randomly inserted into the host chromosome. In about 1–10% of the integrated genomes the lysogenic cycle is established. In these bacteria, the phage DNA remains tightly associated with the host genome and is inherited by daughter bacteria in the course of cell division. The immune state, that is the prevention of superinfection by other Mu particles, is established in the lysogenic bacteria.

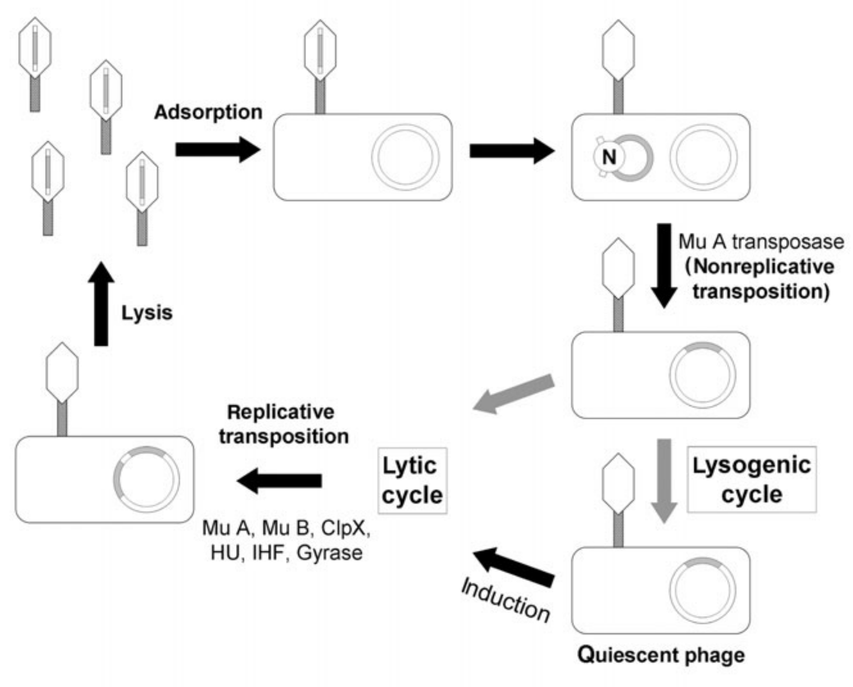


Figure 30-1 Phage Mu life cycle. Redrawn from Sokolsky and Baker (217).

In the majority of bacteria the integrated Mu DNA molecules carry on with their lytic cycle. The protein-A transposase, together with other factors, such as the Mu B protein and various host DNA-replication proteins, amplify these molecules, which are never free in the cytoplasm but instead are transposed to other sites of the bacterial genome. Besides copies of Mu DNA, capsid proteins are produced and the phage DNA is packaged by a "headful" mechanism into the Mu virion particle (see chapter 6). At the end of this cycle, which in *Escherichia coli* needs about 55 minutes at 37 °C, each infected cell produces about 100 phage particles, which are released by cell lysis. These particles are heterogeneous for one of two possible sets of tail fibers, which determine the host range for the next infection cycle and which are determined by the orientation of an internal, invertible region (G-segment, also known as G-region or G-loop). Much of this chapter is devoted to examining the Mu life cycle in increasing molecular detail and references to the mechanisms overviewed in this and the previous paragraph will be provided there.

The Mu Virion

Bacteriophage Mu virions consist of an isometric icosahedral head, 54 nm in diameter, a knob-like neck, a contractile tail, a baseplate and six short tail fibers (2, 239). The DNA isolated from mature phage particles is linear, double-stranded DNA with a size initially estimated to be between 37 and 42 kb (25, 51, 170, 240). This mean value of about 39 kb is represented by the 36,717 bp of Mu DNA present in each particle (185) and 0.5–3 kb of attached, variable-sequence host DNA. The density of Mu DNA is 1.71 g/cm^3, which reflects a G-C content of 52.05% (169, 184, 239).

Electron microscopic analysis of Mu DNA isolated from virions after denaturation and reannealing, restriction fragment mapping, and DNA sequencing (9, 49, 130) reveals three distinct regions of packaged Mu chromosomes: the α-segment, the G-segment, and the β-end. The α-segment is the majority of the genome—the part that is conventionally drawn to the left of the G-segment.The presence of a short (50–150 bp) host DNA segment at the left (*c*) end of the α-segment (9) was demonstrated by Southern blots (27). The G-segment is an invertible region of the Mu DNA that determines host range. The β-end is a short region of the genome to the right of the G-segment that contains the overlapping *com* and *mom* genes. Attached to the right end of the β-region is 1–2 kb of bacterial DNA representing the site where the Mu DNA was previously integrated. When Mu virion DNA is denatured and renatured, this DNA appears as single-stranded "split ends." The presence of this DNA within Mu virions is a consequence of (i) the headful packaging mechanism employed by Mu, (ii) incorporation into virions of only Mu DNA that has inserted into the bacterial chromosome, and (iii) a total Mu DNA length that is substantially less than that which may be theoretically packaged into the heads of Mu virions. The virus-encoded and genome-circularizing N protein is also packaged into the phage head. Both DNA and N protein are injected into the host cell upon adsorption (75).

Genetic Map, Physical Maps, and DNA Sequence

By convention, the left and right extremities of the Mu genome are those that contain the *c* gene and the *mom* gene, respectively (figure 30-2). Sites *att*L (also known as L) and *att*R (also known as R) are also found at the left and right ends of the Mu DNA (figure 30-2) and define the junctions between the phage and the random host DNA sequences (120, 131). Between 1970 and 1985, 36 genetic structures, including genes, regulation sites, and *att* sites, were genetically identified by various groups. The genetic map is the same for the prophage and mature Mu DNA. This contrasts with phage λ (chapter 27), which has a different integration mechanism and consequently different chromosome ends depending on whether the phage DNA is packaged within the virion particle versus or into the bacterial chromosome (1, 9, 168, 192, 263). The maximum recombination frequency for Mu is about 1% (41, 262, 263, 266).

The genetic organization of phage Mu is, as for other bacteriophages, constituted by functional modules. For Mu these modules, whose locations in the genome are indicated in figure 30-2, include (i) an integration-replicative module, from “*att*L” to the “transposition” region, (ii) a module consisting of semi-essential genes (the SEE region), (iii) a morphogenesis module (including a region involved in programmed DNA inversion), and (iv) a module consisting of the *com*–*mom* genes. Gene distribution is roughly colinear with that of the prophage form of bacteriophage λ (246) (see “Mu and Mu-like Phage Evolution,” below). However, due to Mu’s double nature as bacteriophage and transposon, the Mu genome possesses a series of modules characteristic of viral-specific functions as well as a replicative module that confers on Mu its peculiar phage–transposon life-style. It is interesting to note that this last module can be aligned with the transposon Tn3 structure (128).

The Integration-Replication Module

The first module, from the left end of the Mu genome, is implicated in the processes of integration-replication of the phage chromosome. This is a DNA fragment of about 5 kb that is formed by the *att*L site to the left, and genes *A* and *B* to the right. In addition, the *att*R site, which is located at the opposite extremity of Mu DNA, is also involved in integration-replication. Genes *A* and *B* are early genes, essential for Mu DNA integration and replication (58, 191, 266). Gene *A* codes for the phage transposase that is required for integration of the Mu genome into the host chromosome as well as for replication of the phage DNA during the lytic cycle (202, for a recent review, see 30; see also “The Mu Transposase,” below). The *B* gene codes for an ATP-dependent DNA-binding protein that is required for efficient transposition (36, 175). In this region, the *c* and *ner* genes are also present, and are both involved in regulation of the lytic and lysogenic state. The *c* gene codes for the repressor protein (Repc) necessary to establish and maintain the lysogenic state and supply the cell with the superinfection immunity (73, 107, 147, 208, 275). The *ner* gene product (the Ner protein) negatively regulates the transcription from the promoters PcM (also known as Pc or Pc-2) and Pe, and seems essential for phage development (78, 79, 249, 266).

The *att* sites, *att*L and *att*R, are two regions located at the left and right extremities of the phage DNA. Each is formed by three sites (not shown in figure 30-2): L1, L2, and L3 at the left end and R1, R2, and R3 at the right end. These sites are recognized and bound by the Mu A-protein transposase. Interestingly, Repc, besides its function as transcriptional regulator, also competes with the transposase for binding to the two *att* sites.

Likewise, the transposase is able to bind two sites of the Pe operator region, O1 and O2, which are also bound by Repc. These two operator sites, O1 and O2, in fact, constitute what

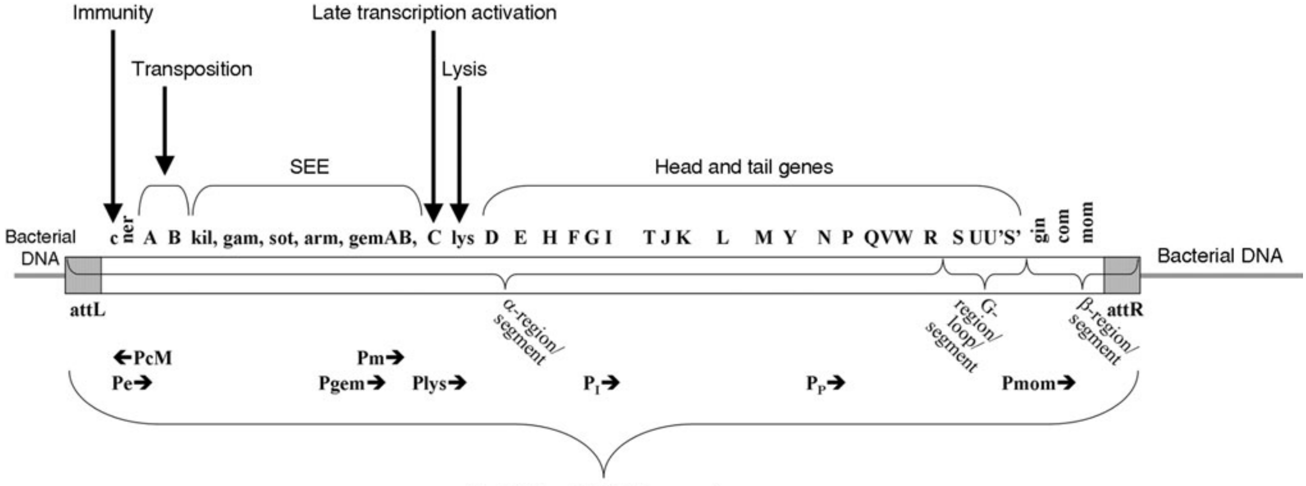


Figure 30-2 Genetic map of the bacteriophage Mu. The arrow tips indicate the transcription starting points.

is described as a transposition *enhancer* (hereafter referred to as "enhancer" or, simply, E) which functions by interacting with the phage *att* sites, *att*L and *att*R (11) (reviewed below under "The Mu Transposase"), and which is also involved in the positive regulation of the transcription of genes *ner*, *A*, and *B*, the latter two products being intimately involved in Mu transposition. The same operator region also plays a critical role in inhibiting transposition by providing the binding target for Repc, which shuts off the transcription of gene *A* and therefore production of its transposase product (80).

Demonstration of the functional autonomy of the integration-replication module is its ability to behave as a transposon if carried by a recombinant plasmid. These plasmids, called miniMu, also harbor a thermosensitive mutation in the repressor gene, *c*ts62. At low temperatures they consequently behave as independent replicons due to Repc activitiy, but can be induced to transpose at the higher temperatures at which the mutant Repc protein is no longer functional (35).

The Semi-essential Genes Module

The region between 4.3 kb and 10 kb of the Mu genome, between the *B* and *C* genes, is described as "semi-essential early" (SEE). Note that this "*C*"gene and the "*c*"gene discussed in the previous section are found at distinct loci. The *c* gene is found outside of and to the left of both the *B* gene (also discussed in the previous section) and the SEE region. The *C* gene, on the other hand, is found at the extreme right end of the SEE region (figure 30-2). Little is known about the biological role of the SEE region although a number of functions which affect burst size, DNA replication, and other aspects of Mu behavior have been identified in it (for review see 195). These functions permitted the identification of various genes such as *kil*, whose expression is lethal for the bacterium, *gam* (by analogy with the *gam* gene of λ), *sot* (stimulation of transfection), *arm* (amplification of Mu replication), and, controversially (i.e., see "Negative and Positive Control of Transcription," below), *cim* (control of immunity).

The SEE region is transcribed as a polycistronic message that originates from the Pe promoter. Within this region, however, another promoter has been identified, the Pgem promoter, which constitutively transcribes the last two genes of the SEE region (55), *gemA* and *gemB* (the latter also known as *mor*), which together are organized as an operon. The GemA protein modulates the expression of various host genes and is responsible for decreasing host DNA gyrase activity and thereby promotes DNA relaxation of the bacterial genome (70, 196). The GemB/Mor protein is involved in Mu late gene transcriptional transactivation (74, 172). As discussed below, under "Lysogenic Conversion," it is likely that both genes impact gene transcription.

The Modules for Morphogenesis

The first two genes of this region (*C* and *lys*) are not directly implicated in morphogenesis. The *C* gene, which defines the right end of the SEE (described above) codes for a positive regulator of late gene transcription while *lys* is involved in bacterial cell lysis and viral particle liberation at the end of the lytic cycle (see chapter 10). Indeed, *lys*$^-$ mutants produce phages that are not released from the bacterial cell. A specific enzymatic activity of Lys has not been shown experimentally (174), though its sequence affinities strongly suggest that it is a member of the true lysozyme family (185).

Genes *D–J*, found immediately to the right of *lys*, are implicated in phage head-protein synthesis (73, 88). However, between the genes *G* and *I*—which is both in the middle of the *D–J* sretch of genes and in the middle of the Mu genome—is also a site that is required for rapid Mu replication (204). This site is efficiently cleaved in vitro and in vivo by *E. coli* DNA gyrase in the presence of enoxacin (203), and has been labeled as the Mu Strong Gyrase Site (SGS). The subsequent genes in this morphogenesis module, *L* through *R*, code for tail proteins. Protein N, coded by gene *N* that is found within this *L* through *R* stretch, is packaged into the phage head and, as noted above, is involved in genome circularization at the start of infection (see "Circularization of Infecting DNA," below). Genes *S*, *U*, *S*′, and *U*′, also found within the morphogenesis module, at its extreme right end, are specifically localized to an invertible sequence called the G-segment (reviewed below under "The Invertible G-Segment").

An additional module consists of the *com–mom* genes. The *mom* gene encodes an unusual DNA modification function that protects the viral genome against a wide variety of host restriction endonucleases (see also "Regulation of Mu Development," below). *Com* is involved in post-transcriptional regulation of *mom* (124, 272).

Promoters and Transcriptional Terminators

The promoter PcM drives the transcription of the repressor gene, *c*, and, to date, is one of only two Mu promoter that drives left-oriented transcription (figure 30-2) (the other, momP2, is discussed below and in reference 229) Various promoters drive rightward transcription, for example Pe for the early region, a middle promoter (Pm), and three promoters for the late transcription, Plys, Pl, and Pp, together with a promoter region called Pmom. The middle promoter, Pm, promotes transcription of the *gemB*/*mor* gene. GemB/Mor activates C transcription and protein C, in turn, activates late gene expression. The other promoters include Plys, Pl, and Pp which control late gene expression (167). Within the region transcribed from Pe, another promoter was localized, Pgem, which constitutively transcribes a small operon, *gem*, that is made up of two genes, *gemA* and *gemB* (reviewed above and below) (17, 23, 98, 141, 142, 205, 207).

An inverted repeat (IR) is found upstream of the *gemA* gene and is centered near the initiation point of Pgem-directed transcription. This IR acts as terminator for the transcript that originates from the Pe promoter (56). Other terminator sequences include IRs that lie downstream from the *gemB/mor* gene (56) and downstream of the *C* gene (104, 165). These sequences, respectively, are found just upstream of the promoters Pm and Plys in figure 30-2.

Mu Genome Sequencing

New data on Mu genetic organization and the identification of new coding regions have been obtained following the total sequencing of phage DNA (185). Fifty-six probable phage Mu genes (open reading frames) were identified using *E. coli* codon usage as a search probe. This number compares to the 36 genes that have been genetically identified. The sequence analysis shows that it was possible in many but not all to correlate the genetically defined genes with open reading frames in the sequence. The majority of the 56 genes have AUG start codons. Three genes—*G*, gene 45 (*Q*?), and *gin*—have GUG start codons, and one (gene *c*, also known as gene 1) is known from an earlier work to use a UUG start (68, 254). The largest gene, encoding the putative tail-length tape measure protein, was not previously identified genetically. It lies between genes *M* and *N* in figure 30-2. The start sites for gene *T*, the major head subunit (which is found, as shown in figure 30-2, between genes *I* and *J*), and gene *L*, the tail sheath subunit, were confirmed by N-terminal sequencing. Many of the predicted Mu genes encode presumptive proteins that show similarity to predicted proteins from a prophage in *Haemophilus influenzae*, as well as to predicted proteins encoded by prophage genes in other bacteria.

Fate of Infecting DNA

Once in the cell, the Mu DNA is circularized via the action of the N protein and, as seen with other temperate bacteriophages, Mu can choose between the lytic and the lysogenic cycle (figure 30-1) (see chapter 8). However, in constrast to the other temperate bacteriophages, for phage Mu DNA integration (also known as transposition) is a prerequisite for both cycles (161). Two transposition mechanisms take part in the Mu life cycle. The first is a nonreplicative (or conservative) transposition that mediates the integration of the infecting phage DNA into the host genome (6, 33, 90, 158). The second mechanism, called replicative transposition, is exploited instead to replicate the integrated phage DNA. In the latter mechanism one daughter Mu genome remains in place and the other daughter is replicated while integrated into a second site of the bacterial genome (35).

Circularization of Infecting DNA

Circularization of infecting DNA within the host cell is a rather common phage strategy (see chapter 7) that is used to protect phage genome termini from nuclease attacks, to convert the linear genome to an integrative precursor, or to represent the replicative form of the genome. Various mechanisms allow the achievement of this objective, such as covalent closure of the phage DNA (single-stranded tails at each 5′ or 3′ end can be complementary and therefore cohesive), recombination between redundant terminal sequences, or by means of a protein bound to the extremities of the genome that converts the linear phage DNA into a noncovalently closed form. Circularization of Mu DNA exploits this last mechanism.

The existence of a phage-encoded protein injected into the bacterium during adsorption that was responsible for this noncovalent closure was suggested by Kahmann et al. (122). They observed that the Mu DNA was scarcely infective in bacteria that had been made competent by treatment with Ca^{2+}. However, infectivity could increase in bacterial hosts lacking exonuclease V. Nevertheless, transfection still was very low compared, for example, with that obtained with λ DNA: about 1000-fold lower. The infectivity could increase by about 2 orders of magnitude, however, if the Mu *sot* gene was expressed in the recipient bacterium. The *sot* gene is also called *gam*, and it may inhibit a host nuclease (7). These facts were explained with the hypothesis that proteins, normally injected with the Mu genome and necessary for a positive outcome of the transfection, could be eliminated during the phenol extraction of Mu DNA (122). This hypothesis was sustained by the fact that the DNA extracted from the viral particles broken by freeze-thaw methods was about 1000 times more infective than DNA extracted with phenol (38). In addition, treatment of DNA made by freezing and thawing with proteinase K or with pancreatic DNase resulted in a loss of infectivity.

Extension of the above hypothesis is that the infective form of Mu DNA is a DNA–protein complex and that the protein portion of this complex consists, at the very least, of Mu protein N. The following observations support this extension: (i) CsCl-purified DNA was found to be noncovalently associated with a 65 kDa polypeptide, but no protein was found in phenol-extracted Mu DNA (38). (ii) A supercoiled form of infecting Mu DNA was isolated by Harshey and Bukhari (93). Upon treatment with pronase, phenol or sodium dodecyl sulfate, however, this supercoiled DNA was converted to a linear, Mu-length form, indicating that the circle was not covalently closed but instead held together by proteinaceous material. (iii) A protein noncovalently bound to Mu DNA was identified in minicells infected with this phage (209). The 64 kDa polypeptide was found to co-sediment with Mu DNA through a sucrose gradient. (iv) Implying that complementary DNA ends are not involved in the interaction, Mu DNA circularization does not require removal of the *E. coli* DNA sequences that are attached to both ends of the Mu genome in the viral particle (209). (v) Antiserum to the protein purified from viral particles is specific for the Mu N product (75).

Conservative Transposition

The term "conservative transposition" (also known as nonreplicative transposition) means that, upon infection, both strands of the Mu DNA double helix are integrated without any previous DNA replication. Three kinds of experiments allowed determination of this mechanism. In the first experiment, Liebart et al. (158) showed that infection of isotopically labeled "heavy" cells containing a small plasmid resulted in the formation of some Mu-containing plasmid DNA of a density consistent with integration without replication. The second, carried out by Akroyd and Symonds (6), found that a cell transfected with an artificially constructed heteroduplex Mu DNA molecule, containing one wild-type strand and one mutant strand, gave rise to a mixed burst with an approximately equal number of wild-type and mutant phage. The interpretation is that both strands of the infecting DNA were integrated but then segregated during subsequent DNA replication. The third experiment, by Harshey (90), showed that infecting phage DNA that was fully methylated (from phage grown on a strain that overproduces the *E. coli dam* methylase) was, when injected into a *dam*-mutant host, integrated in a fully methylated form.

Lysogenic and Lytic Cycles

The integrated Mu DNA is committed toward two alternative cycles: lysis or lysogeny. Only in a minority of cases is the lysogenic cycle established, whereas the lytic (or productive) cycle is predominant in most infections.

Frequency of Lysogenization

Under conditions in which all the cells of a bacterial population are infected (high multiplicity of infection), the Mu lysogenization frequency, amongst survivors, varies between 1% and 10%(108). More stringent determinations have been obtained with the Mu mutant *c*ts Amp in which the ampicillin-resistant gene is inserted into the G-segment (154) (see figure 30-2). The frequency of ampicillin-resistant bacteria, obtained upon infection with this phage and due to its stable integration (lysogenization), is on the order of 0.1% per phage particle.

Superinfection Immunity

In the lysogenic bacterium the constitutive expression of the *c* gene product, Repc, makes the cell immune to Mu superinfection. This state is stable and therefore the frequency of spontaneous prophage induction is very low, about 10^{-4} (159). Neither chemical nor physical factors are known to induce the prophage. However, as has also been observed for phage λ, prophage transfer by conjugative mating induces a lytic cycle in the recipient bacterium (zygotic induction) that results in cell death. The zygotic induction can be prevented, however, by using a Mu-immune recipient bacterium. On the other hand, the Mu *vir* mutant, which carries a mutation in the *c* repressor gene (68, 253), kills a Mu lysogen upon infection but with a mechanism that is totally different from the one used by λ vir (see "The Mu Repressor," below). Experimentally, induction of lysogenic bacteria is obtained using phages carrying a *ts* mutation in the repressor gene, such as Mu *c*ts62 (107), which is the most popular. These lysogens are stable at 30 °C and are induced at 42 °C.

Random Integration of Mu DNA

In lysogenic bacteria, Mu DNA is inserted into randomly distributed sites. About 2–3% of lysogens carry auxotrophic mutations due to prophage integration in genes implicated in the corresponding biosynthetic pathways (234). Consistent with the randomness of this integration, Bukhari and Zipser (26) describe the mapping of 75 independent insertions of Mu DNA in the *lacZ* gene alone, showing that each integration event occurred in a different site. From these facts we derive the classical notion of Mu random integration. Many other investigations reinforce these observations (47, 50; for a review see 91). However, the random character of Mu integration does not exclude the existence of integration hot spots, such as the *malK–lamB* region of *E. coli* (210, 225). Other factors can also modify integration frequency. It has been observed, for example, that Mu integration in *lacZ* is reduced when that gene is actively transcribed (60). Adding to the randomness of Mu intregration, the Mu genome's insertion can occur in either of the two possible orientations of the phage genome (109). Furthermore, analysis of the bacterial DNA flanking the inserted Mu DNA revealed the existence of 5 bp direct repeats of the target-site DNA. These 5 bp vary from one site to another and are a consequence of the Mu integration mechanism (30).

Proteins implicated in Mu lysogenization are those required for phage DNA integration. These phage-coded proteins are Mu A (the phage transposase) and Mu B, which stimulates integration frequency. Gene *A* amber mutants do not lysogenize and do not integrate their DNA into the bacterial host, as shown by the behavior of ^{32}P-labeled gene *A* amber mutants (34, 161). Less evident is the role of Mu B in the process. Mu *B* amber mutants show an approximately 3-fold reduction in lysogenization frequency when compared with the wild-type phage (190). However, these mutants are at least 100-fold defective for replication (37, 39), indicating a different impact of Mu B on integration than on replication.

Lytic Replication

The lytic cycle is the destiny of most integrated Mu DNA molecules. This cycle can be studied either in infected bacteria or, more easily, upon induction of lysogenic bacteria carrying a thermosensitive mutation in the repressor gene. Even though Mu DNA replication cannot be separated from a general model of transposition, in this section we will examine only the main characteristics of the replication process, whereas analysis of the transposition mechanism will be treated in the next section (for a recent review see 31).

Mu DNA replication starts between 6 and 8 minutes after induction (255, 264) and continues for about 40 minutes. Replication is semiconservative (201) and only two phage-encoded proteins are implicated: Mu A and Mu B. In this process, the integrity of the Mu terminal sequences is essential (246, 256). During the lytic cycle, various rearrangements, such as deletions, inversions, replicon fusions, and transpositions are produced in the host cell. These rearrangements are a consequence of the transpositive replication itself. In addition, free Mu DNA molecules are never found in the cell, although circular forms of DNA consisting of Mu DNA associated with heterogeneous sequences of bacterial genome can be observed.

The critical experiment showing that Mu DNA replication is peculiar compared with that of other bacteriophages was performed by Ljungquist and Bukhari (159). These authors showed that upon induction of a Mu lysogenic strain, Mu DNA replication occurs in situ without excision of the prophage from the bacterial chromosome. Restriction endonuclease digestion, which cuts both bacterial and phage sequences, and the separation on agarose gel of the fragments obtained, showed DNA fragments containing both phage and bacterial DNA. These fragments span the junction between the prophage extremities and the bacteria DNA. The persistence during the entire lytic cycle of the junction fragment seen in the lysogen suggested that the Mu DNA was replicated in situ, that is without leaving the bacterial genome. In addition, during the lytic cycle other junction fragments between Mu and bacterial DNA of various sizes accumulate, providing evidence that in situ replicated DNA is integrated into new sites of the genome.

Packaging of Mu DNA

Mu DNA packaging is an oriented process starting with the recognition of the *pac* site (87), which resides 32–54 bp to the right of the left end of the Mu chromosome, between the transposase binding sites L1 and L2 found within *att*L (89). The cut in the DNA at the left end occurs in the host DNA flanking the Mu insertion site at about 100–200 bp to the left of the *pac* site. Therefore, 50–150 bp of this bacterial sequence are encapsidated in the phage particle. The packaging mechanism is "headful," as deduced by Bukhari and Taylor (25) who observed that insertions in Mu DNA were compensated by a reduction in the bacterial DNA length that is associated with the Mu DNA. That is, the total length of encapsidated DNA depends on the size of the phage head and, since more than the 37 kb of Mu DNA can be inserted in the phage capsid, the additional DNA consists of the bacterial DNA flanking the integrated Mu genome, with most of that DNA flanking the right end of the Mu genome.

Little is known about Mu packaging at the molecular level. It is still not known what determines the *pac* specificity, the cut specificity to the left of the *c* gene, or the proteins implicated in *pac*-site recognition. Although mutants in genes *E* and *I* do not make DNA molecules of mature Mu length, there is no proof of a direct role of their gene products in Mu packaging. Partially addressing this dearth of knowledge, an in vitro system for maturation and encapsidation of the Mu-like phage D108 was set up by Burns et al. (28). In this assay a crude extract of cells in a late phase of Mu lytic cycle was able to mature and encapsidate the D108 DNA starting from the Mu packaging machinery. The assay revealed that ATP is necessary for the reaction and that the D108 transposase inhibits this process. Replication is not required for either Mu or phage D108 DNA packaging.

Conservative and Replicative Transposition

A peculiar characteristic of Mu bacteriophage is, as already stated, its ability to exploit two alterative transposition pathways, which characterize different stages of its life cycle. A nonreplicative mechanism characterizes the integrative transposition that mediates the phage DNA integration in the host genome (also known as conservative transposition as well as integrative transposition), whereas replicative transposition allows the generation of new copies of the viral genome ultimately found within the phage burst at the end of the lytic growth. In this section we consider the molecular mechanisms of Mu transposition in greater detail.

Transposition

The transposition mechanism is formally described in the Shapiro model (220). A staggered cut of the target DNA sequence is made at two specific phosphodiester bonds found on opposite DNA strands, one located adjacent to each of the two phage DNA ends (L1 and R1) (figure 30-3). The resulting 3′-OH groups are joined via transesterification to two phosphodiesters placed 5 bp apart on the two strands of the target DNA. This is a process known as the

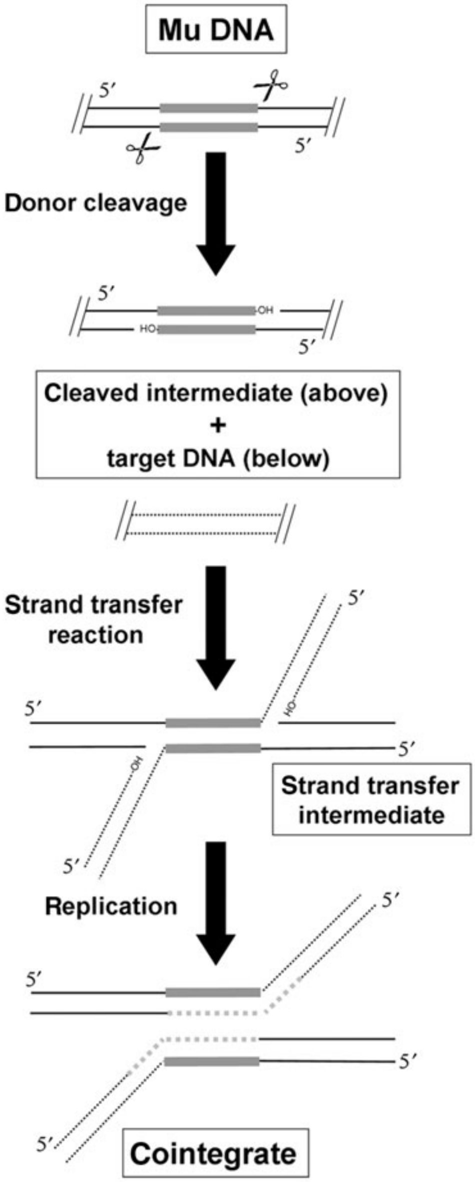


Figure 30-3 The Shapiro model for transposition.

"strand transfer reaction" which gives rise to what is described either as a "strand transfer" or "Shapiro" intermediate. The free 3′-OH sites, generated between the host and the Mu DNA, are needed to initiate continuous-strand replication. Covalent extension of these ends, generated during transposition, would explain why DNA polymerization occurs only toward and then into the Mu DNA rather than also outward into the bacterial chromosome. This replicative process, which is mediated by host proteins, yields a transposition product in which each phage end is covalently attached to both the flanking DNA and the target DNA. Subsequent separation of the DNA strands of the phage double helix, as occurs during semiconservative DNA replication, ultimately separates the donor and target DNA molecules.

The transposition process was reconstructed in vitro by Mizuuchi using partially purified proteins (179). The requirements of that system are a supercoiled miniMu substrate, which contains the left end (*att*L) and the right end (*att*R) transposase binding regions and the O1 and O2 enhancer elements. Also required are the Mu protein-A transposase, the Mu B protein, two host-encoded accessory proteins (HU and IHF), and a target DNA molecule into which miniMu integration will occur (153, 180). The Mu nucleoprotein (also known as transpososome) complex consists of all these elements and its composition is now well characterized (for a review see 31).

Conservative Transposition

It is still poorly understood how Mu performs nonreplicative (also known as conservative or integrative) transposition. Phage DNA enters the bacterial cell in a linear form and is circularized by the N protein (75), but the role of this protein–DNA complex on integrative transposition has not been determined. In addition, both strands of phage DNA, but not the host sequence flanking the infecting DNA, are incorporated into the new host. Early in vitro studies of Mu transposition suggested that the outcomes of the two transposition pathways result from alternative processing of the Shapiro intermediate (43, 44), that is conservative transposition could originate from the repair of the strand transfer intermediate, coordinated with deletion of the bacterial DNA carried by the pre-integration circularized phage genome, to generate a simple insertion.

Recently the analysis of Mu *B* mutants—which are able to stimulate integration to a comparable extent as wild-type Mu *B* but are unable to support the formation of the Shapiro intermediate during in vivo replication—suggested that nonreplicative and replicative transposition in phage Mu may diverge before formation of the Shapiro intermediate (216). In addition, using both gyrase-inhibiting drugs and gyrase mutants, Sokolsky and Baker (226) showed, as consistent with previous studies, that gyrase inhibition causes severe defects in replicative transposition since complexes between the Mu protein-A transposase and Mu DNA, and thus transposomes, fail to form. By contrast, gyrase activity is not essential for conservative transposition (226).

Replicative Transposition

Contrary to what happens during integrative transposition, recombination steps during replicative transposition are well characterized. The initial step for transposition requires the assembly of higher order protein–DNA complexes, the transpososomes, in which the two phage DNA extremities, a transpositional enhancer (i.e., the O1 and O2 binding sites) also called IAS (internal activator sequence), which is reviewed below, and a tetrameric Mu protein-A transposase take part. In addition, a number of protein cofactors

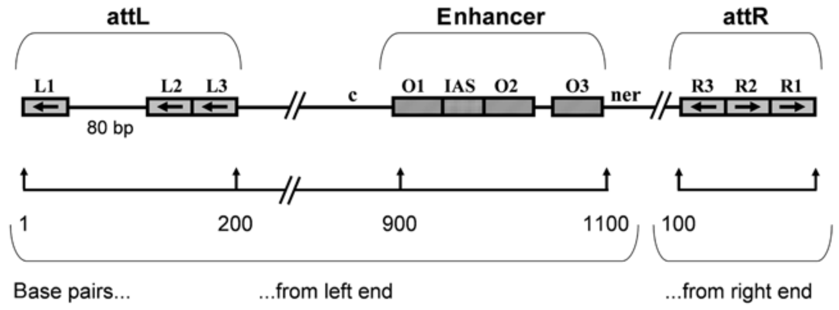


Figure 30-4 The transposase targets. The *att* site, located at the phage extremities, and the enhancer sequence are recognized and bound by the phage transposase. Each *att* site is formed by three binding sites for Mu A (see ''Replicative Transposition''). These structures are also recognized by the phage repressor, which shows overlapping binding specificity with the transposase.

participate, including the Mu protein B, which serves as both ATPase and target DNA activator.

The *att* sites, to which the protein-A transposase binds, reside at the two extremities of the phage genome (figure 30-4). Each *att* site, *att*L and *att*R, is formed by three transposase-binding sites, L1, L2, L3 and R1, R2, and R3 (40). These three binding sites are related by a 22 bp consensus sequence, 5′-GTTTCAYNNRAARYRCGAAAR(A/C), that shows no obvious internal symmetry (46). Binding of the protein-A transposase to these sites results in bending of the Mu end DNA by approximately 60–90° (3, 53). The three binding sites of *att*L are all oriented in the same direction but have a different spacing: L1 and L2 sites are separated by about 80 nucleotides while L2 and L3 are rather contiguous. On the contrary, the three *att*R sites are adjacent, but the R3 site is oriented in the opposite direction with respect to R1 and R2. The binding affinity of the protein-A transposase also appears to be greater for L1, L3, R3 sites, as determined by DNaseI footprinting of the A protein on the linear ends. By contrast, R1 and L2 are weak and R2 intermediate (11, 150, 276). On the other end, footprinting of the core type1 complex (see "Transposition intermediates") showed that A protein covers sites L1, R1, and R2 (150).

Sites O1 and O2 of the operator, to which the Mu Repc protein binds, contain two clusters of Mu protein-A binding sequences (IAS sequence) (181). The binding affinity of the A-protein transposase for IAS is lower than the affinity of the Mu A protein for the two *att* sites, *att*L and *att*R (156). The *att* sites, however, are implicated in a complex circuit of interactions with the transpositional enhancer sequences, O1 and O2 (11, 184), as discussed below (particularly under, "The Transposition Intermediates").

There are two important aspects to the transpositional enhancer's functionality. The first is the integrity of the whole IAS region. In fact, when the two hemisites are separated by means of digestion with restriction endonuclease, the stimulating effect is no longer observed even if a high concentration of each of the two segments, O1 and O2, is present. The second aspect that characterizes the transpositional enhancer's functionality is the DNA bending due to binding of IHF (a host-encoded accessory protein) to the site located between O1 and O2 (215). Actually, in vitro experiments with miniMu showed that the stimulating effect of the transpositional enhancer is mainly observed when complexes between the L and R extremities of the Mu chromosome (*att*L and *att*R, respectively) are formed, while the assembly of complexes either with a couple of R-ends or with one of the L-ends does not account for the presence or absence of the transpositional enhancer. Therefore, the transpositional enhancer helps to avoid an incorrect pairing of Mu-extremity L- and R-ends (259) and its function could be to stabilize the transition state from what is known as a three-site synapse—between the L-end, the Enhancer, and the R-end—which is called an LER complex (182).

This model is supported by the observation that the transpositional enhancer becomes dispensable once the transpososome is formed (183, 231). However, the transpositional enhancer remains associated with the *att* sites even after strand transfer has been completed (as shown in the type 2 complex presented on the far right of figure 30-6). Perhaps the sequestration of the transpositional enhancer within the LER complex prevents the Mu Repc protein from binding to the enhancer, thus signaling a commitment to transposition (201).

The Mu Transposase

The Mu A transposase is a complex protein with a modular organization (figure 30-5). In solution it is a monomer of 663 amino acids with a molecular weight of 75 kDa (94) and is able to bind target sites on phage DNA. However, it is in a tetrameric form that it promotes the DNA cleavage and joining reactions of transposition (reviewed in 31, 152, 180). All six *att* sites (L1–L3 at the left end and R1–R3 at the right end) and the two operator sites, O1 and O2 (which constitute the transposition enhancer), are bound by the Mu A protein in its monomeric form. The Mu A tetramer, by contrast, binds only to three of the *att* sites: L1, R1, and R2. Partial proteolysis allowed the identification of the various protein-A domains (189). The N-terminal domain is responsible for DNA binding and contains motifs for recognition of the two types of DNA sites bound by the transposase: the phage extremities (*att*L and *att*R) and the enhancer site (O1 and O2). In particular, domain Iα binds the enhancer site and domain Iβγ binds the two *att* sites.

The central domain, domain II, contains in its N-terminal proximal part (subdomain IIα, which is not explicitly shown in figure 30-5) three amino acid residues: Asp, Asp, and Glu (DDE) at positions 269, 336, 392, respectively. These residues are essential in the strand-cleavage and strand-transfer transposition steps and constitute the catalytic core of the transposase enzyme (14, 136, 143). Mechanistically, the three residues' function is coordination of divalent metal ions necessary for catalysis (162, 268). Subdomain IIβ (also not shown explicitly in figure 30-5) has a large positive charge potential (214) and has been implicated in metal-assisted assembly of the protein-A transposase tetramer.

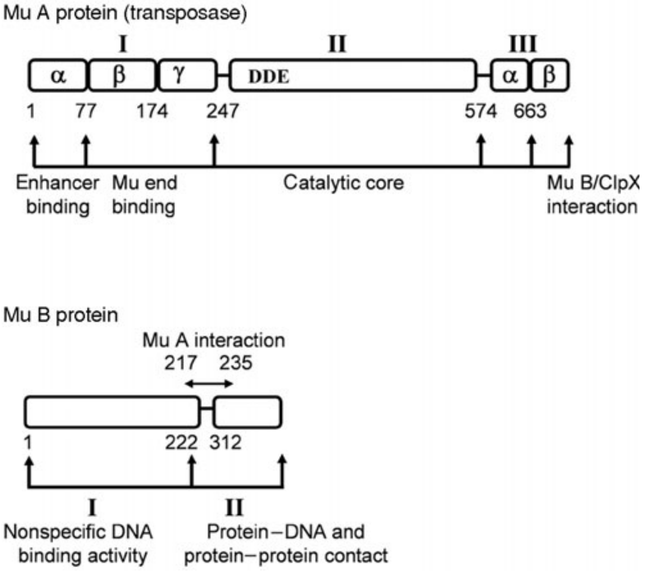

Mu A protein (transposase)
I
II
III
α
β
γ
DDE
α
β
1
77
174
247
574
663
Enhancer binding
Mu end binding
Catalytic core
Mu B/ClpX interaction
Mu B protein
Mu A interaction
217 235
1
222
312
I
II
Nonspecific DNA binding activity
Protein–DNA and protein–protein contact

Figure 30-5 Schematic representation of MuA and B proteins.

The C-terminal domain, or domain III, is also required for assembly of the transposase tetramer and probably for its chemical competence. Domain IIIα has both nonspecific DNA binding and cryptic nuclease activity (16, 155, 271). Domain IIIβ interacts with the Mu B protein, promoting strand transfer to the target DNA (15, 155, 270). Domain IIIβ also interacts with the host-encoded chaperone, ClpX (157), which is involved in disassembly of the transpososome after transposition and prior to replication (145).

The Role of Mu B

Mu B, a protein of 312 amino acids (figure 30-5), is an ATP-dependent DNA binding protein that is able both to capture the target DNA and to interact with the other complexes of the transpososome (186). Mu B acts as an allosteric activator of the Mu A transposase, although the molecular mechanism by which Mu B stimulates the transposase is unknown. Williams and Baker (267) showed that Mu B activates transposition by stimulating the reaction step between cleavage and joining that is otherwise slowed by the 3′-flanking strand, that is the bacterial DNA located between the integrated Mu DNA ends and the sites of DNA cleavage (these 3′ ends are shown in figure 30-3 immediately above the label "Cleaved intermediate").

Mu B is formed by two globular domains, of which the N-terminal domain (25 kDa), which contains what is known as an AAA$^+$ ATPase motif (189), shares structural similarity with the N-terminal domain of the *E. coli* helicase, DnaB (112), and shows nonspecific DNA binding activity (figure 30-5). The 11 kDa C-terminal domain (237) plays an important role in the protein–DNA and protein–protein contact needed to regulate transposition (40). Peptide binding experiments revealed that the region of Mu B that binds to Mu A lies between residues 217 and 235, a segment included in the N-terminal part of the domain II. In vivo,

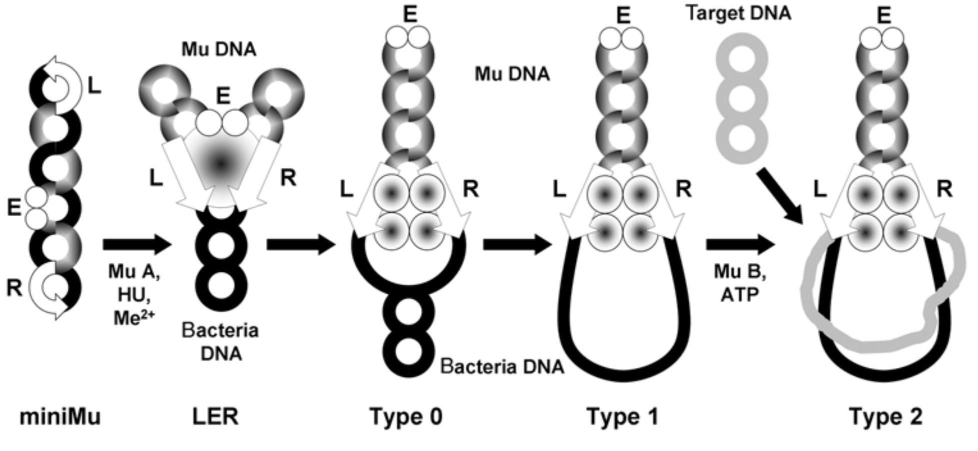

L
E
R
Mu DNA
E
L
R
Mu A, HU, Me2+
Bacteria DNA
E
Mu DNA
L
R
Bacteria DNA
E
L
R
Target DNA
E
Mu B, ATP
L
R
miniMu
LER
Type 0
Type 1
Type 2

Figure 30-6 Transpososome assembly during Mu replicative transposition. Redrawn after Jiang et al. (113). The process is described in the text, under "Replicative Transposition."

proteins lacking residues 295–312 are able to promote integrative transposition but, interestingly, not replicative transposition (34, 216).

Mu B is also involved in the process called “target immunity,” namely the inactivation of DNA sites proximal to the transposon as potential transposition targets and a phenomenon shared by many transposable elements (42). Mu uses a mechanism that favors its insertion in sites located 10–25 kb from the original location of the phage genome (164), protecting an entire supercoiled domain of the bacterial chromosome. This allows Mu to avoid having its genome inactivated by an intramolecular DNA strand reaction. Although the process has not completely elucidated, a key step is the formation of Mu B oligomeric complexes, which are formed by the protein binding AT-rich DNA regions (3, 83, 84). Through studies employing total internal reflection microscopy, which permits examination of Mu protein B behavior on a single DNA molecule, Greene and Mizuuchi (85) showed that, after a stochastic nucleation event, assembly of Mu B oligomeric complexes occurs by a mechanism involving the sequential binding of small units of MuB to form large polymeric complexes. The Mu A-protein transposase then stimulates the Mu B ATPase function, causing dissociation of the Mu B protein from the DNA. One model predicts formation of a multivalent complex between the tetrameric form of the Mu A protein and the Mu B oligomer, which thereby catalyzes processive disassembly of Mu B from the potential DNA target (83, 84). Consequently, Mu B accumulates only in those DNA regions not bound by Mu A (3, 83, 84). The DNA decorated with Mu B oligomers is then recognized by Mu as a potential target for its genome insertion, whereas the naked DNA is not recognized (15). Thus, Mu B dissociation from transpososome proximal sites makes these DNA regions inaccessible as possible targets for the DNA strand transfer reaction.

The Transposition Intermediates

Initiation of phage Mu DNA transposition requires assembly of higher order protein–DNA complexes called transpososomes as identified by in vitro studies (31, 44, 179). In the presence of a target-DNA substrate such as a supercoiled miniMu plasmid, the phage-encoded A protein, bacterial products Hu and IHF, and an appropriate divalent ion, the LER initial complex is formed (figure 30-6). In this structure, through complex protein–protein and protein–DNA interactions, the Mu left (L/*att*L) and right (R/*att*R) ends and the enhancer element (E) are gathered (259). By analyzing the events preceding LER complex formation, Pathania et al. (201) showed that through Mu A-protein transposase the transposition enhancer interacts solely with *att*R and that this event precedes the entry of *att*L into the synaptic complex.

LER is an unstable complex that, in the presence of Ca^{2+}, is quickly converted into the first stable complex, the transpososome type 0 (183, 258), where the two Mu ends are engaged within the active site of the transposase (figure 30-6). In this structure the enhancer is no longer associated with the ends and Mu A has tetramerized (32). In the presence of magnesium, the 3′ ends of Mu DNA are nicked and the Type 0 complex is converted into a type 1 complex, also known as a cleaved donor complex (45, 232). The two subunits within the Mu A tetramer, which are associated with sites L1 and R1 (sites that undergo cleavage), provide in trans their DDE amino acid residues towards the strand cleavage/transfer reaction.

The addition of MuB, ATP, and target DNA results in the formation of type 2 complex (figure 30-6). This is also known as a strand transfer complex (STC) and is the product of the transfer of the 3′ ends of the Mu DNA to a target DNA molecule (32, 45, 232). In the absence of target DNA, Mu B and ATP stimulate intramolecular strand transfer, with the 3′ ends of Mu DNA transferred into a new DNA site of the same donor molecule (15, 175).

The Transpososome-Replisome Transition

Nakai et al. (188) describe an assay based on that of Mizuuchi (179 and described above) that reproduces in vitro the transition from transpososome formation (as discussed immediately above) to the replication of Mu DNA, which occurs during the Mu lytic cycle. The transition is thought to resemble homologous recombination with strand exchange giving rise to DNA replication. The strand exchange step requires only five components: Mu protein-A transposase, the Mu B protein, HU (a host factor), supercoiled Mu DNA, and a target DNA substrate. A total of eight purified *E. coli* replication proteins are then employed, along with the Mu strand exchange product (labeled “strand transfer intermediate” in figure 30-3), to effect replication of the Mu chromosome in the course of integration into the target DNA substrate.

It has been observed that the Mu A protein (the transposase) remains so tightly bound to the DNA that it blocks the action of the *E. coli* replication proteins necessary to replicate the strand transfer intermediate (187). After removal of Mu A, these *E. coli* proteins can replicate the strand transfer intermediate, forming what is known as a cointegrate (figure 30-6) that consists of two daughter Mu genomes, one integrated in the original (donor) position and the second integrated at a second (target) position (or DNA). The Mu B protein and the *E. coli* HU protein have also been found to be loosely bound to this DNA making up the strand transfer intermediate (151). Therefore, the type 2 (strand transfer) complex (STC) must be destabilized before replication can occur (157). In fact, after phenol extraction of the transpososome, some of the replication proteins can enter the replicative forks: under these conditions the E. coli DNA pol I catalyzes a limited strand displacement synthesis (116, 144).

STC conversion to the cointegrate occurs if, besides the eight proteins, a partially purified fraction of the host enzymes (Mu replication factor, MRF) (144) is added. However, no cointegrates are formed when *E. coli* DNA Pol III, DnaB, and DnaC are omitted from the reaction system. Therefore, the transpososome bound to the template imposes a strict requirement for both MRF and the specific replication proteins in order to initiate Mu DNA synthesis. The MRF complex was thereafter separated into two fractions, MRFα and MRFβ, each formed by many components (187). MRFα includes the chaperone protein, ClpX, and other as yet unidentified factors called MRFα2.

The *E. coli* ClpX protein is involved in reactivating damaged proteins (224) and can associate with *E. coli* ClpP protein (a serine peptidase) to form the ClpXP protease complex that is involved in protein degradation (86). The ClpXP complex degrades the Repc (the Mu repressor) and thereby induces the lytic cycle. It also is required for in vivo Mu replication (177). This protein is responsible for the first steps in the transition from transpososome to replisome. The ClpX protein recognizes a specific peptidic sequence at the A-protein transposase C-terminal extremity (domain III, figure 30-5) and, through its unfolding activity, remodels the strand transfer complex (at this point, STC1) to form another more fragile complex, called STC2 (145). In this complex, the phage extremities are still maintained in a synaptic state. The Mu A protein is activated so that it can recruit crucial host factors needed to initiate Mu DNA synthesis by specific replication enzymes (145). Together with the unidentified host factors known as MRFα2, the transpososome's role is the pre-primosome assembly at the level of Mu DNA forks. MRFβ consists of the PriA, PriB, and DnaT proteins that make up the *E. coli* primosome (115). Its role in vivo was confirmed by the fact that Mu cannot give rise to its lytic cycle or replicate its DNA by transposition in knockout *priA* (115) or *dnaT* (218) mutants. It was hypothesized that these primosome components promote the primosome assembly for phage Mu at the site of homologous strand exchange (188).

The transition steps from transpososome to replisome could be: (i) The molecular chaperone ClpX converts STC1 to STC2 (types 1 and 2 in figure 30-6), altering the conformation of the transpososome. (ii) MRFα2 then displaces the transpososome to assemble the prereplisome at the Mu forks to form what is known as strand transfer complex 3 (STC3). (iii) PriA, a component of both MRFβ and the *E. coli* primosome, binds to the forked DNA structure created by strand exchange (the strand transfer intermediate) and begins the assembly process of a replisome at one Mu end. PriA promotes the assembly of the replisome, preferably starting from the left end as occurs during in vivo Mu replication. The mechanism that determines which Mu end is used to initiate DNA synthesis is not yet clear. (iv) PriA then assembles a prereplisome complex by recruiting *E. coli* PriB, DnaT, and DnaB–DnaC complexes. In this process, DnaB must be bound to single-stranded lagging-strand template—which is not to be confused with the "lagging strand" observed in the formation of Okazaki fragments in systems that possess replication forks. To create this binding site, PriA unwinds duplex DNA by translocating 3′ to 5′ along this template. Once bound to the DNA, it attracts primase to form a primosome which catalyzes primer synthesis for lagging strand synthesis. Meanwhile, *E. coli* DnaB promotes binding of the DNA pol III holoenzyme to complete replisome assembly (188).

Regulation of Mu Development

Upon infection of a sensitive host or induction of a lysogenic bacterium, Mu DNA is transcribed by the bacterial RNA polymerase in a cascade event. This transcription is strongly asymmetric and unbalanced in the rightward direction since only the repressor gene, *c*, is leftward transcribed (i.e., from the promoter, Pc, also known as Pc-2 or PcM; figure 30-2). Transcription of bacteriophage Mu overall can be divided into three phases: early, middle, and late.

Early Transcription

Early transcription starts about 1.5 minutes after induction of a Mu lysogenic strain and lasts about 6 minutes. Hybridization studies have indicated that, during this phase, this transcription is directed rightward on the conventional Mu map (13, 250, 265). Transcription begins from Pe (141) and transcribes the early region (13, 250, 265) from 1 to about 8 kb from the left end of Mu DNA (56) (see figure 30-2). In this region we find the essential genes, *A* and *B*, which are implicated in Mu integration and replicative transposition, and the *ner* gene whose product negatively regulates the amount of early transcript during the lytic cycle (245, 249, 266). Soon after, the transcription from Pe enters the semi-essential region. The *kil* gene is immediately transcribed. The other genes of the region (77)—with the exclusion of the *gem* operon, which is submitted to its own regulation (56)—are transcribed after a delay that is perhaps due to a polymerase pausing. Neither Mu protein synthesis nor DNA replication is required for early transcription (169, 265, 266).

Middle Transcription

Middle transcription starts 4–8 minutes after induction and increases until bacterial lysis. S1 mapping experiments localized the start site of this transcript about 740 bp upstream of the *C* gene, which is under the control of the middle promoter, Pm (227, 228). The sequence analysis of the Pm DNA region shows a significant similarity with the −10 consensus sequence of *E. coli* promoters, but the homology with the −35 region is weak (227). The middle transcription differs from early and late transcription since it needs both Mu protein synthesis and Mu DNA replication.

The Mu middle promoter is positively regulated by one of the *gem* operon gene products, GemB/Mor (74, 172). By studying the kinetics of Mu transcription after induction of a Mu *c*ts lysogenic strain that additionally carries an amber mutation in the *gemB* gene, Giusti et al. (74) observed that the pattern of early transcription (i.e., the first 10 minutes) was essentially identical to that of a *gemB*$^+$ strain. However, after the *ner*-dependent pause, the *gemB*$^-$ induced lysogen displayed a dramatic alteration in transcription abundance: the expression of late genes was both slowed down and reduced in the mutant with respect to that of *gemB*$^+$ phage. In addition, the β-galactosidase activity produced by a plasmid in which the *C* gene promoter, Pm, was fused with the *lac*Z gene was stimulated after infection with Mu wild-type, but not after infection with the *gemB*$^-$ mutant. Analogous results were obtained by Mathee and Howe (172) who observed that the β-galactosidase synthesis expressed by a plasmid carrying the fusion Pm-*lac*Z increases more than 20 times in bacterial strains where the Mu prophage is induced.

How GemB/Mor acts is poorly understood. Mathee and Howe (173) showed that Mor is not an alternative σ subunit; Kahmeyer-Gabbe and Howe (126) hypothesized that it could function as an accessory DNA-binding protein for activation of *E. coli* RNA polymerase. Mor binds the Pm promoter in a sequence between −56 and −33 from the transcription starting point and this region is also bound by the phage repressor, Repc. It is therefore possible to hypothesize that the repressor might negatively regulate middle transcription (126). Since GemB/Mor is produced, at a low level, along with GemA in lysogens (56), it might activate Pm and produce enough C protein to activate the phage morphogenesis and lysis functions and thereby reduce cell viability. It was therefore hypothesized (126) that Mu may have evolved additional safeguards to prevent expression of *C* and subsequent lytic functions in a lysogen. One of those safeguards could be the direct repression of Pm by Repc. A second safeguard could be the requirement for Mu DNA replication for Pm activity.

Late Transcription

Late transcription starts about 10–12 minutes after induction and increases for 45–60 minutes until bacterial lysis. Three late promoters, Plys, Pl, and Pp, together with the Pmom region (23, 166, 167), are involved in this transcription to express the genes responsible for capsid synthesis, Mu DNA modification, and cell lysis. The C product acts as an activator for the RNA polymerase necessary for this transcription (23, 99, 166, 167, 228). The *C* gene codes for a 16.5 kDa polypeptide (140 amino acids) that is a site-specific DNA-binding protein (17, 23, 71, 210).

C protein could act as an accessory factor of the RNA polymerase, like the cI or cII protein of λ or the host proteins CAP or OmpR. These and other activators bind DNA at or upstream of the −35 regions of specific promoters allowing recognition and activation of the host RNA polymerase (166). In the four late promoters the A and T residues in the second and sixth positions, respectively, of the consensus hexamer (TATAAT) for the σ^{70} subunit of the bacterial holoenzyme RNA polymerase are conserved, but the −35 sequence (TTGACA) is substituted by the sequence (ccATAAcCcCPuG/Cac). This sequence renders late transcription C-dependent (167).

Mom Transcription

The modification function expressed by genes *com*–*mom* produces α-*N*-(9-β-D-2′-deoxyribofuranosylpurin-6-yl)glycinamide (95, 233). The adenine residues to be modified form part of the deduced consensus sequence c/gAc/gNPy (119). The *mom* gene belongs to a dicistronic operon located at the right extremity of the Mu genome whose two genes (*com* and *mom*) partially overlap (for reviews see 97, 121). The premature expression of *mom*, as that of *lys*, is harmful for the host cell and therefore expression of these genes is strictly controlled (98, 125). In the case of *mom*, the regulation acts at both transcriptional and translational levels (97).

The Pmom promoter region (figure 30-2) contains the approximate location of two promoters: momP1 and momP2. The promoter momP1 is relatively weak, with a consensus sequence of −35 (ACCACA) and −10 (TAAGAT), separated by 19 bp containing a run of six T nucleotides. RNA polymerase does not bind the momP1 promoter (17); instead it binds weakly to the diverging momP2 promoter, which overlaps momP1 and promotes "leftward" transcription (71, 229). The stretch of six A nucleotides complementary to the run of six T nucleotides (above) appears to be part of a UP element, a promoter region found upstream to a −35 region that assists in the promotion of leftward transcription from momP2 (230).

At the transcriptional level, *mom* expression is activated by the C protein, which binds the sites at −28 and −57 of the momP1 promoter region (71, 211) and brings about an asymmetric distortion and unwinding of the DNA (18, 19, 211). It has been suggested that the leftward transcription of momP2 might prevent low-level rightward transcription from momP1 in two possible ways. The first is that momP2 might compete with momP1 for RNA polymerase binding in the absence of C protein; the second is that leftward transcription produces an antisense transcript that might prevent *gin* mRNA elongation into *mom*. The first hypothesis was discarded when the destruction of the −10 hexamer of momP2 did not lead to an increase in the basal-level activity of momP1 hexamer (19). These facts have been interpreted with the idea that momP2 acts as a sink for capturing RNA polymerase in the vicinity of momP1, so that RNA polymerase is ready for occupancy at momP1 when the regulated expression of C protein has accumulated sufficient protein levels to turn on *mom* expression (17).

The Com protein acts as a positive regulator (119) at the level of translation of the *mom* open reading frame (100, 273). Com appears to relieve A translational repression caused by the presence of a TIS (translation-inhibitor stem-loop) structure (101, 271, 274). Thus, footprinting studies demonstrated that Com specifically bound to *com–mom* mRNA at the putative TIS site destabilized the messenger RNA secondary structure TO expose the ribosome-binding sequence and the Mom start codon (101, 274).

Negative and Positive Control of Transcription

Phage Mu can enter into a lysogenic cycle only when transcription of the Mu early region is repressed by the *c* gene product, Repc. By contrast, entry into the Mu lytic cycle occurs when gene *c* expression is repressed and, as a consequence, early genes are transcribed. After an initial early burst of transcription 4 minutes after virion adsorption or prophage induction, the transcription rate progressively decreases due to expression of the *ner* gene, the first gene of the early region. Early transcription then continues at a low level during the lytic cycle (265). The Ner protein inhibits the transcription from PcM, the promoter controlling gene *c* expression (249). Mutant analysis evidenced that the Ner binding site overlaps Pe and PcM (249) and footprint analysis with the purified protein localized the Ner binding site between sites Pe and PcM. This site shows a dyad symmetry and binding of Ner to the left or right half may inhibit transcription from Pe or PcM, respectively (79).

There is no evidence for the existence of a positive control of Mu repressor synthesis. Contrasting λ bacteriophage, there are no Mu genes equivalent to the λ *cII* and *cIII* genes and all Mu clear-plaque mutants belong to the same complementation group, mapping to gene *c*. The identification of a particular phenotype in Mu *c*ts lysogens, defective for replication and unable to show recovery in immunity when shifted from 42 to 32 °C (72), lead to suggestion of the existence of a gene for immunity control, called *cim*. This behavior can only be observed in a kil^- background, though it is now thought that this immunity effect is not due to the action of a separate gene, but instead reflects a subtle interplay occurring between various elements that control early gene expression from the early Mu region. Various data also indicate that some host functions, such as IHF, may stimulate early Mu transcription (80).

The Mu Repressor

The 22 kDa Mu repressor (Repc) is a product of gene *c* and consists of 196 amino acids. Its function is the repression of Mu lytic growth and in doing so Repc binds to a total of 11 sites on the Mu chromosome: nine sites spread among three enhancer-region operators (O1, O2, and O3 as presented in figure 30-4) and the two promoters, Pe and PcM. The early promoter Pe overlaps the O2 operator site and the early transcription (through O2 and O3) of genes *A* and *B* that are necessary for lytic growth. The promoter PcM, on the other hand, is co-localized with O3 and drives transcription in the opposite direction of Pe, through O2 and O1 and into gene *c* (for a review see 80). Cooperative binding of Repc to both O1 and O2 is thought to inhibit transcription at the Pe promoter and therefore to inhibit both early transcription and lytic growth. Cooperative binding to all three operator segments (O1, O2, and O3) at higher Repc concentrations also shuts down the PcM promoter and therefore synthesis of Repc (253). The repressor also inhibits transposition directly by competing for Mu protein-A (transposase) binding to the enhancer which is located within the operator, between O1 and O2, and part of the LER complex required for transposition (above) (4, 47, 156).

Two kinds of *c* mutants have been isolated: thermosensitive mutants and virulent mutants. The Mu *c*ts62 (109) mutant, employed to create temperature-inducible Mu lysogens or miniMu plasmids (as discussed above), carries the substitution R47Q found in the N-terminal Repc DNA binding domain. This mutation reduces repressor binding to DNA at 30 °C and makes it very weak at 42 °C. Alternatively, a virulent Mu mutant that has frameshift mutations altering the last 11–26 C-terminal residues of Repc (68) can superinfect Mu lysogens to disrupt Mu immunity and induce lytic development (253). The alteration causes the mutant repressor (Vir) to be highly sensitive to ClpXP protease (discussed above). This Vir mutant confers a dominant negative phenotype because it also promotes rapid degradation of the wild-type Repc repressor, even though by itself wild-type Repc is otherwise relatively resistant, in vivo, to ClpXP proteolysis (67, 148, 177, 260).

Mu Ner

The first gene transcribed from Pe is *ner*, whose protein negatively regulates both *c* (249) and the transcription of the early region (10, 266). Ner is a small (74 amino acid), basic DNA binding protein and, on the basis of its function, can be grouped with the proteins regulating the lytic–lysogenic switch, such as λ Cro (for review see 177). Despite this fact, there is no sequence homology between Mu Ner and Cro proteins or other DNA binding proteins of other phages, except the Ner protein expressed by the Mu-like phage D108 and the *E. coli* Nlp protein (229). Constitutive *ner* expression (due to an exogenous promoter carried on a multicopy plasmid) shuts off early transcription by infecting phage and also inhibits Mu replication (78, 249). This phenotype, similar to Repc-mediated immunity, is called pseudo-immunity.

The essential character of the Mu Ner protein is proved by the behavior of ner^- mutants, which greatly affect the

Mu lytic cycle but not Mu lysogeny, which remains normal. The phenotype of the double mutants *ner*$^-$*c*$^-$ helped the comprehension of the Ner mechanism of action (79). These mutants not only do not lysogenize but also do not form plaques unless they are complemented for Ner. This result suggested th in the absence of Ner$^-$ phage growth is not due to an excess of repressor synthesis but instead could be explained assuming that the RNA polymerase, during *c* gene transcription, proceeds through the protein-A transposase binding sites found in the vicinity of gene *c* (figure 30-4), thereby preventing transposase binding (79). In other words, since Ner appears to be essential even in the presence of an inactive repressor, then a reasonable hypothesis is that it is repressor gene transcription and not the repressor protein itself that interferes with transposition.

Commitment to Lysogeny versus the Lytic Cycle

The integrated Mu DNA is committed to two alternative cycles: the lytic (or productive) cycle, which concerns most of the population, and the lysogenic cycle. It is not known what determines this choice, though a distinct possibility stems from the fact that the early transcripts from the Pe and PcM promoters overlap so that the transcription of one interferes with the transcription of the other. Thus, the choice between the lytic and lysogenic state is dependent, at least in part, on the balanced use of the two promoters. In addition, it has been hypothesized that the transposition enhancer is a structural element critical for the lysis/lysogeny choice. This structure, which includes the operator sites of the phage (figure 30-4), is recognized and bound by both the Repc repressor protein and the protein-A transposase, which stems from their sharing sequence homology in their N-terminal domains. Thus, depending on which of the two proteins is bound, the enhancer/operator can bring about two mutually opposite outcomes: turning off phage early gene expression or turning on transposition (201).

Lysogenic State Derepression

The lysogenic state is remarkably stable and spontaneous prophage induction very low. The lysogenic state seems to be maintained by cooperative binding of the Repc repressor to both O1 and O2, inhibiting transcription at the Pe promoter and thereby preventing expression of the Mu *A* and *B* genes needed to effect lytic growth. Nevertheless, the repressor, as regulator of the Mu transposition, must allow immune-state derepression, presumably in response to some change in host physiological state. Conditions and signals that lead to the prophage derepression, however, are unknown. A series of data show that lysogens, with a Mu *c*ts prophage, can be derepressed during the stationary phase or upon carbon starvation (S derepression). This is a process that requires the ATP-dependent proteases ClpXP and Lon as well as the stationary-phase-specific sigma factor, σ^s (149, 212, 223).

Even if the primary signal that leads to derepression is not known, lysogen immunity is influenced by the activity of many host-derived factors besides the ClpXP protease (177, 221) including IHF (8, 64, 142, 222, 251, 252), FIS (20, 21), H-NS (62, 194), DNA binding proteins (81), and the RpoS/σ^s stationary phase sigma factor (76). From these observations it emerges that Repc is a sensitive receptor of degradation signals, apparently becoming susceptible to protease degradation in response to one or more *E. coli* signal molecules. Such a mechanism was initially suggested as the means by which proteolysis can be triggered by the mutant form of Repc called Vir, which can cause wild-type repressor to be degraded at an accelerated rate by the *E. coli* ClpXP protease (67, 148, 177, 260). In addition, derepression can be promoted by repressor peptides that include the N-terminal binding domain (DBD) and the C-terminal *ssrA* tag of Repc (193). The latter is an 11-residue degradation signal (AANDENYALAA) added to incomplete peptides by *ssrA* RNA (tmRNA) when ribosomes stall on mRNA (82, 134, 244). The modulation of Repc degradation of the ClpXP protease may, therefore, be a mechanism by which entry into lytic development is regulated: Repc binds to the Mu operator sequences, O1, O2, and O3, which regulate expression of the transposition functions as well as the repressor gene (46, 142) (as discussed above).

Mu Excision

Despite its behavior as a transposable element, mutations due to the integration of Mu DNA are remarkably stable: reversion frequencies are less than 10^{-10} (117). This lack of reversion of Mu-induced mutations is paradoxical when compared with the behavior of other transposable elements (both prokaryotic and eukaryotic), all of which are able to excise precisely. However, lysogens with the Mu wild-type prophage, submitted to particular conditions such as long periods without cell division or nutritional starvation, or lysogens carrying Mu prophage with a *gem* operon mutation, show relatively frequent excision.

The first observed Mu DNA excision was performed by Bukhari (24) with bacterial strains lysogenic for a phage mutant (Mu *c*ts62 dX) that are able to express the transposase after induction of the lytic cycle by thermal shift without either host cell killing or prophage induction. Under these conditions, both precise and imprecise excision events were obtained, all of which were associated with prophage loss. Mu *c*ts62 dX excision required partial derepression of the Mu *A* gene, coding for the transposase, and unlike excision by other prokaryotic transposons was RecA-dependent (135). For example, Tn*10* excision is independent of RecA (65, 163) and Tn5 excision is similar to spontaneous deletion events (55), but without any involvement of the transposase.

An interesting kind of Mu prophage excision was observed with the Mu *gem*2ts mutants. Mutations induced by the integration of a Mu *gem*2ts mutant prophage can revert at frequencies around 1×10^{-6}, more than 10^4-fold higher than that obtained with Mu wild-type. Several aspects characterize Mu *gem*2ts precise excision (69): (i) the phage transposase is not involved in the excision process but is necessary for phage reintegration (P. Ghelardini, personal communication); (ii) the RecA protein is not necessary; and (iii) revertants remain lysogenic with the prophage inserted elsewhere in the host genome. The site of reintegration somehow depends on the original site of insertion. There is a strong correlation between the original site of insertion (the donor site) and the target of the phage DNA migration (the receptor site) (197).

The Invertible G-Segment

Host Range Modification

An aspect of Mu biology that has raised much interest is its ability to extend its host range through the programmed inversion of a segment of its genome (figure 30-7). The first evidence of this inversion was the observation that Mu DNA obtained after induction of a Mu *c*ts lysogen, after melting and reannealing, showed a non-pairing bubble in 50% of the molecules due to the inversion of a DNA segment of about 3000 bp (49, 51 and, for review, 139) called the G-segment. This observation suggested that in half the phage particles the G-segment had the opposite orientation compared with the other half of the population. However, the phage produced through the lytic cycle in *E. coli* K12 was in 98% of cases in an orientation conventionally indicated as G(+) and only in the remaining cases G(−) (132). Symonds and Coelho (234), with a single burst experiment, suggested that only the G(+) particles are able to infect *E. coli* K12 strains, whereas the G(−) cannot grow in these strains. In addition, this experiment permitted determination that the inversion is a slow process, since many generations are necessary to observe a G(+) cell starting from a single cell lysogenic for a G(−) prophage.

The inversion occurs in the prophage and an equilibrium between G(+) and G(−) cells is reached in a culture that has increased in size (via lysogen division) to more than 1×10^7 cells. Thus, a Mu lysogen population by and large is a mixed population consisting of 50% G(+) and 50% G(−) prophages. Only the G(+) particles of this population can be propagated by infection in *E. coli* K12. The G(−) particles, by contrast, can infect other bacterial species including *Citrobacter freundii* (247). Various groups (139) have also shown that in the G(−) configuration Mu can infect *E. coli* C, *Shigella sonneei*, *Enterobacter*, and *Erwinia*. Meanwhile, while in the G(+) state, besides *E.coli* K12, Mu can infect *Salmonella arizonae* (127). The receptor of both G(+) and G(−) particles is in the membrane lipolysaccharides. No differences were observed by electron microscopy in the tail fibers but antibodies obtained against Mu (G+) particles are more active against these particles than against those with an opposite orientation, and vice versa (88).

The product of the *gin* gene, adjacent to the G-segment, mediates the inversion (132). S^- and U^- mutants do not adsorb on *E. coli* and lack tail fibers (88). These two genes coding for the proteins involved in adsorption are located in the left portion of the G-segment, in the G(+) orientation. In the G(−) orientation the *S*′ and *U*′ genes codify for an alternative set of tail fiber proteins.

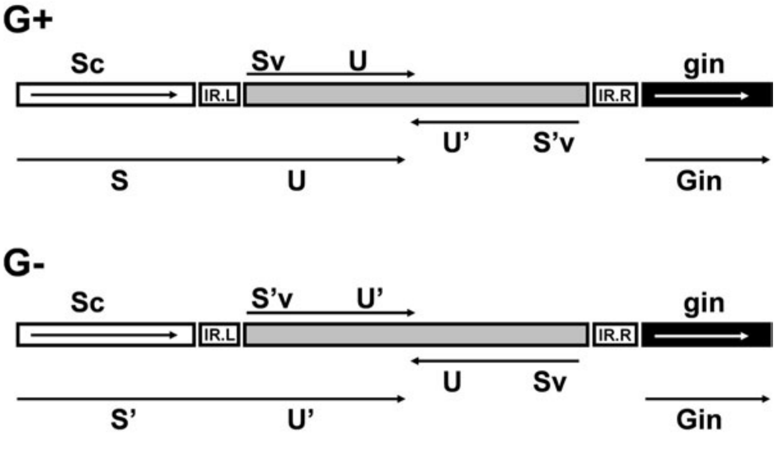


Figure 30-7 G-region inversion. The alternative expression of *S U* and *S* ′*U*′ genes, due to the programmed inversion of the 3 kb DNA segment between the IR sites, allows the synthesis of an alternative set of tail fiber proteins. The *S* gene of bacteriophage Mu is formed by an Sc sequence common to both *S* and *S*′ genes, localized before the IR.L site, and a variable portion, Sv or S′v, localized within the invertible G-region and followed by the *U* and *U*′ genes, respectively.

The Inversion Mechanism

Various very sensitive biological assays were developed to study the G-segment inversion mechanism (124, 206). In one of these, the *lac* operon of *Escherichia coli* is inserted in the G-segment fused with the 5′ extremity of the Mu *S*′ gene (figure 30-8). Bacterial Δlac cells, harboring the plasmid in which this insert was cloned, are not able to ferment lactose. If the plasmid is incubated with bacterial extracts overproducing Gin, the G-segment is inverted, the *S*′ gene is expressed, and the hybrid product S′-β-galactosidase is produced. Bacterial cells transformed with this plasmid are able to grow in minimal lactose medium. In another system, the Kn^R lacking its own promoter is cloned in the G segment carried by a plasmid and, as predicted, inversion switches the harboring bacterium from Kn^S to Kn^R.

Inversion, which is RecA-independent, was concluded to be mediated by the 21.7 kDa Gin protein (132, 133, 205) which catalyzes an intramolecular recombination between the two 34 bp inverted repeats at the extremities of the G-segment (205). Gin is an invertase, a family of proteins able to invert DNA segments, which are highly homologous (more than 65% of identical amino acids) and therefore able to complement each other (129). On this basis it is not surprising that their target sites are also homologous among themselves. The invertible region length, on the other hand, is not critical: deletions reducing the G-segment from 3 kb to 132 bp do not alter the inversion frequency. On the contrary, the target sites orientation is very important: if they are colinear instead of inverted, the inversion reaction efficiency decreases by about two orders of magnitude.

Inversion of the G-segment occurs by *gin*-mediated recombination between 34 bp inverted repeats located at the ends of the G-segment (111, 123, 131, 219). The inverted repeats (IR.R and IR.L), which comprise the reaction target sites, constitute two Gin-invertase binding sites. Each inverted repeat is formed by two hemisites, which are not equivalent to each other: site I can be bound even if present alone, whereas site II can be bound only if site I is already occupied by the protein, suggesting an invertase cooperative binding. In any case, the recombination efficiency is maximal when both sites are present. It has furthermore has been observed that a DNA region on the right of IR.R is essential for high-frequency G inversion (124). This region, called *sis*, acts as an enhancer since neither its orientation nor its distance from the IR sequences substantially modifies the stimulation effect which, however, is observed only in *cis*. The *sis* region is constituted by a 60 bp sequence formed by a triple repetition of a consensus sequence binding site for the bacterial protein, Fis (138). The function of Fis (140) is to form a DNA–protein complex that helps to correctly assemble the synaptic complex where the crossing-over reaction occurs.

Additional Considerations

Lysogenic Conversion

Bacteriophages, as well as other genetic elements such as plasmids and transposons, besides coding for essential functions for their maintenance in the host cell can carry genes that modify seemingly unrelated aspects of the host cell phenotype (lysogenic conversion; see chapter 27). In the case of Mu bacteriophage, the lysogenic cell shows changes in expression of a high number of bacterial genes due to constitutive *gemA* expression. Genes affected include some DNA replication and cell division determinants (70, 196). This modification in host gene expression is accompanied by both a reduction in the supercoiling of the bacterial DNA and a reduction in bacterial gyrase activity. A model proposed by Ghelardini et al. (70) correlates these facts with hypotheses that the Mu-induced relaxation of the bacterial chromosome should modify the transcription of these genes since the expression of many bacterial genes is controlled by supercoiling of their respective promoter (for review see 54). Bacterial cell-cycle reprograming, furthermore, is observed upon infection with some *gem* mutants or in bacterial cells where *gemA* is highly expressed.

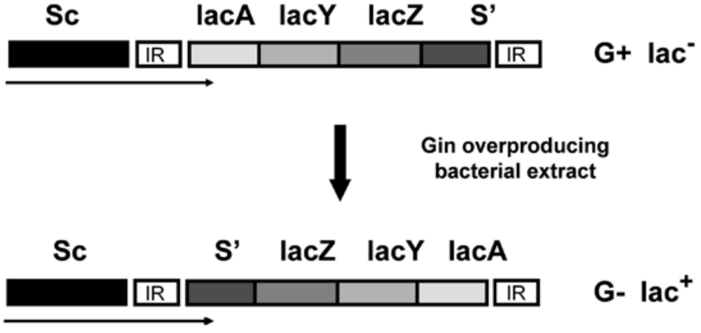


Figure 30-8 Biological assay for the G-loop inversion. The *E. coli lac* operon, inserted in the G-loop fused with 5′ extremity of the Mu *S*′ gene, is inverted after incubation with a bacterial extract overproducing Gin. The inversion allows *lac* operon expression and renders the bacterial cells able to grow in minimal lactose medium.

The importance of these facts in Mu biology is still poorly understood. However, it is interesting to note that, recently, genes homologous to *gemA* have been identified in the *Haemophilus influenzae* Mu-like prophage FluMu (63), in *Neisseria meningitidis* prophages Pnm1 (137) and Pnm2 (200), and in *E. coli* O157:H7 strain Satai prophage Sp18 (13) (these Mu-like phage strains are reviewed below). It has been hypothesized that *gemA* could belong to the "morons" family of transcriptional elements (118, 185) which constist of a σ^{70} promoter, a coding region, and a factor-independent terminator, all located between two genes whose homologs are adjacent in a different phage (see chapter 27). These elements could be independent modules which have entered only recently into the phage genome (106).

Mu as a Genetic Tool

Random integration into the host genome, induction of strong polar mutations as a consequence of its integration, and induction of various types of genome rearrangements (inversions, deletions, duplications of the adjacent genes, replicon fusions, and host DNA segment transposition) are properties of Mu biology that, since their identification, at the basis of the development of in vivo genetic engineering. The development of these techniques started as soon as it was possible to separate the host genome rearrangements, during the Mu lytic cycle, from the phage lethal effects. This was accomplished through the construction of a large collection of clever Mu derivatives. These techniques were applied to gene manipulation not only in *E. coli*, the bacterial host where Mu was identified, but also in many Enterobacteria, as *Shigella*, *Salmonella* (*S. typhimurium*, *S. typhi*, *S. montevideo*), *Serratia marcescens*, *Citrobacter freundii*, *Enterobacter* sp. *Klebsiella pneumoniae*, and *Erwinia* sp. Mu has also been found to be effective manipulating in the Rhizobiaceae, *Agrobacterium tumefaciens*, various species of the *Rhizobium genus*, in the *Pseudomonas genus*, and in a score of additional Gram-negative bacteria (for reviews, see 243, 248).

Other Mu-Like Prophages

For a long time, Mu was the only transposable bacteriophage isolated, while λ- and P22-like phages were independently isolated many times (52, 113). The inexplicability of this fact raised the question of whether Mu is a vanishing breed which Larry Taylor saved from extinction (128). In reality, a heteroimmune Mu phage (D108) was identified in 1971 and, thereafter, two other transposable phages were isolated, one infecting *Pseudomonas* and the other *Vibrio cholerae*. Recently, completed genome sequences have permitted identification of Mu-like prophages in *Haemophilus*, *Neisseria*, and *Deinococcus* (185). A comparison between these various Mu-like bacteriophages (105) is given below.

Coliphage D108

Phage D108 is a transposable bacteriophage isolated by Mise (178) that is highly correlated with Mu. The two phages share 90% homology as determined by DNA hybridization. As for phage Mu, phage D108 presents variable ends of bacterial origin at its genome extremities, an invertible region corresponding to the G-segment, and generates a 5 bp duplication in the integration target site. In addition, anti-Mu antibodies cross-react with phage D108 virions and all the genome rearrangements induced by Mu phage transposition are also observed with phage D108.

Non-homologous regions between Mu and D108 correspond to those of the repressor and *ner* genes and at the 5′ end of the *A* gene. The consensus sequence for transposase binding also shows some similarities. Functional hybrid phages containing the Mu left end (genes *c*, *ner*, and the 5′-end *A*, the transposase gene) and the remaining D108 genome were constructed. These phages were called MD phages whereas those with the D108 left end and the remaining Mu genome were called DM phages (242).

Pseudomonas Transposing Phages

A number of transposable bacteriophages were isolated in *Pseudomonas* strains by Krylov and colleagues (146). They all show host sequences at their genome extremities, their genome being about 37 kb in length. Among them, D3112-like phages (146, 217) and B3-like phages (4, 5) constitute the two main groups of *P. aeruginosa* transposable phages. Phage D3112 shows a genome structure and a life cycle strictly correlated to that of Mu (12, 22, 185, 213, 217) whereas the virion structure is morphologically more similar to that of λ-like phages, and for this reason D3112 is classified as a member of the λ-like *Siphoviridae* family (phages with long, noncontractile tails) (see chapter 2 for a primer on phage classification by virion morphology). Phage D3112, therefore, resembles a kind of hybrid with a Mu-like genome organization and a λ-like tail morphology (22). Upon infection, the D3112 phage DNA integrates in the host genome by a transposition mechanism and, in this state, the Repc protein represses the lytic cycle and stabilizes integration (217). The D3112 complete genome sequence (257) is 37,611 bp in length and encodes 53 potential open reading frames, including three known genes (the repressor *c* gene and early genes, *A* and *B*). Forty-eight percent of open reading frames were similar to Mu-like phages and prophage sequence, including the proteins responsible for transposition, transcriptional regulation, and virion. However, phages D3112 and Mu do not share sequence homology (213).

Vibrio cholerae Mu-like Phages

Little is known about VcA1 and VcA2 phages, isolated from *V. cholerae* strain NIH41 (198). These two phages are heteroimmune to each other and show only small morphological differences (66). However, it has been observed that VcA1 can integrate in various sites on the host genome and that its properties are similar to those of the Mu phage (114).

Haemophilus influenzae prophage FluMu

The *H. influenzae* Rd genome, between bases 1,559,722 and 1,594,398, carries a Mu-like prophage called FluMu (63). This prophage, 34,676 bp in length, shows a relatively high G-C content: about 50%, compared with the 30% of the host genome. The FluMu genome has terminal 5′-TG-3′ dinucleotides flanked by five direct repeats of bases (ACGCA) of the host DNA. Three potential transposase binding sites were identified at each end of the phage DNA (185). A comparison between the FluMu and Mu genomes reveals a colinear genetic organization. In the early region, only genes *ner*, *A*, *B*, *gam*, and *gemA* have significant matches at the amino acid level. Contrarily to Mu, FluMu appears not to have an invertible region and indeed a gene product homologous to Mu invertase has not been identified in FluMu.

Neisseria meningitidis Mu-like Prophages

Two genomes (type A and type B) of *N. meningitidis* have been completely sequenced (199, 238). A prophage, called Pnm1 (137), and three others (171), highly related to phage Mu, were identified in type A *N. meningitidis* bacteria. The Pmn1 G-C content is 53.1% versus the 51.8% of the host bacterium and the phage genetic organization shows many characteristics in common with the Mu-like element.

Deinococcus radiodurans R1 Mu-like Prophage

D. radiodurans R1 (261) carries a Mu-like prophage called RadMu whose most interesting aspect is the high phylogenetic distance between *Deinocossus* and the proteobacteria host of Mu, FluMu, and Pnm1 (269).

Mu-like Prophage Sp18 in E. coli O157:H7 Strain Satai

Sequence analysis of the *E. coli* O157:H7 strain Satai (103) revealed the presence of a Mu-like prophage called Sp18 whose genetic organization is a kind of mosaic very similar to the other Mu-like phages (185).

Mu and Mu-like Phage Evolution

One of the characteristics of the genetic organization of bacteriophages is the clustering of genes that belong together functionally. This idea was expressed for the first time by Campbell and Botstein (29) to explain the organization of the λ genome, which can be subdivided into different functional modules. Making a comparison between the organization of bacteriophage genomes is a way to group them into families and to compare the single representative of each family with those of another family and to state their phylogenetic relationships. Of course, other relationships can be searched for, such as DNA, protein, or functional homologies or similar genome-replication strategies. In the case of Mu, the transpositive module also constitutes the replicative and the integrative module. By observing this structure, one may question the evolutionary interrelationships between phages and transposons. For example, comparison of the Mu and λ prophage genetic maps shows a similar and colinear organization of the genes (246):

Mu : *c* *ner* *A B* SEE region
C lysis genes structural genes
λ : *cI* *cro* *O P* b2 region
Q lysis genes structural genes

Comparison with the transposon structure suggests that the Mu replicative module is similarly comparable with the Tn3 transposon family. Mu has the same type of replicative transposition as the Tn3 family, including 5 bp duplication at the insertion site. Also, the transposase and the repressor gene are transcribed in conflict. According to Kamp (128), Mu could have originated from the insertion of a Tn3-like transposon into the genome of a Mu progenitor, due to the mobilization of phage modules between two inverted copies of the transposon. One transposon copy at the left of this composite transposon could have retained most of its functions, while the other could have lost its transposase and had its TnpR resolvase changed into the Gin invertase. Analogously, the origin of Tn3 from a Mu deletion, as suggested by the miniMu behavior, cannot be excluded.

Perspectives

Four decades of research on Mu bacteriophage are characterized by a number of extraordinary discoveries crossing over the various fields of basic biology. A retrospective analysis of the history of this research shows that the transposition mechanism has been the main object of study, particularly in recent years, since this process can be finely dissected in vitro. From a strictly virological point of view, however, many aspects of Mu biology remain unresolved and not adequately studied. For example: the mechanism leading to the integration of the infecting DNA, the role and molecular nature of the products from the semi-essential region, the conflict between commitment to the lysogenic versus productive cycles, the excision mechanism and its biological significance, and lysogenic conversion.

Indeed, more attention needs to be paid to these aspects since they could hide many fascinating features about this virus—a virus that has chosen transposition as a way of life.

Acknowledgments

We thank G. Di Lallo for stimulating discussions and M. Lo Ponte for careful revision of the text.

References

1. Abelson, J., W. Boram, A. I. Bukhari, M. Faelen, M. Howe, M. Metlay, A. L. Taylor, A. Toussaint, P. Van de Putte, G. C. Westmaas, and C. A. Wijffelman. 1973. Summary of the genetic mapping of prophage Mu. Virology. 54:90–92.
2. Admiraal, G., and J. E. Mellema. 1976. The structure of the contractile sheath of bacteriophage Mu. J. Ultrastruct. Res. 56:48–64.
3. Adzuma, K., and K. Mizuuchi. 1991. Steady-state kinetic analysis of ATP hydrolysis by the B protein of bacteriophage mu. Involvement of protein oligomerization in the ATPase cycle. J. Biol. Chem. 266:6159–6167.
4. Akhverdian, V. Z., E. A. Khrenova, M. A. Reulets, T. V. Gerasimova, and V. N. Krylov. 1985. Characteristics of phages-transposons of *Pseudomonas aeruginosa* belonging to 2 groups distinguished by DNA–DNA homology. Genetika 21:735–747.
5. Akhverdian, V. Z., E. A. Khrenova, V. G. Bogush, T. V. Gerasimova, and N. B. Kirsanov. 1984. Wide distribution of transposable phages in natural *Pseudomonas aeruginosa* populations. Genetika 20:1612–1619.
6. Akroyd, J. E., and N. Symonds. 1983. Evidence for a conservative pathway of transposition of bacteriophage Mu. Nature 303:84–86.
7. Akroyd, J. E., and N. Symonds. 1986. Localization of the *gam* gene of bacteriophage Mu and characterization of the gene product. Gene 49:273–282.
8. Alazard, R., M. Betermier, and M. Chandler. 1992. *Escherichia coli* integration host factor stabilizes bacteriophage Mu repressor interactions with operator DNA in vitro. Mol. Microbiol. 6:1707–1714.
9. Allet, B., and A. I. Bukhari. 1975. Analysis of bacteriophage mu and lambda–mu hybrid DNAs by specific endonucleases. J. Mol. Biol. 92:529–540.
10. Allet, B., M. Payton, R. J. Mattaliano, A. M. Gronenborn, G. M. Clore, P. T. Wingfield. G. Kukolj, P. P. Tolenhancer, and M. S. DuBow. 1989. Commentary on the article: Purification and characterization of the Ner repressor of bacteriophage Mu. (FEBS Lett. 244: 369–375). FEBS Lett. 251:282.
11. Allison, R. G., and G. Chaconas. 1992. Role of the A protein-binding sites in the in vitro transposition of mu DNA. A complex circuit of interactions involving the mu ends and the transpositional enhancer. J. Biol. Chem. 267: 19963–19970.
12. Autexier, C., and M. S. DuBow. 1992. The *Escherichia coli* Mu/D108 phage ner homologue gene (*nlp*) is transcribed and evolutionarily conserved among the Enterobacteriaceae. Gene. 114:13–18.
13. Bade, E. G. 1972. Asymmetric transcription of bacteriophage Mu-1. J. Virol. 10:1205–1207.
14. Baker, T. A., and L. Luo. 1994. Identification of residues in the Mu transposase essential for catalysis. Proc. Natl. Acad. Sci. USA 91:6654–6658.
15. Baker, T. A., M. Mizuuchi, and K. Mizuuchi. 1991. MuB protein allosterically activates strand transfer by the transposase of phage Mu. Cell 65:1003–1013.
16. Baker, T. A., M. Mizuuchi, H. Savilahti, and K. Mizuuchi. 1993. Division of labor among monomers within the Mu transposase tetramer. Cell 74:723–733.
17. Balke, V., V. Nagaraja, T. Gindlesperger, and S. Hattman. 1992. Functionally distinct RNA polymerase binding sites in the phage Mu mom promoter region. Nucleic Acids Res. 20:2777–2784.
18. Basak, S., and V. Nagaraja. 1998. Transcriptional activator C protein-mediated unwinding of DNA as a possible mechanism for *mom* gene activation. J. Mol. Biol. 284: 893–902.
19. Basak, S., L. Olsen, S. Hattman, and V. Nagaraja. 2001. Intrinsic DNA distortion of the bacteriophage Mu momP1 promoter is a negative regulator of its transcription. A novel mode of regulation of toxic gene expression. J. Biol. Chem. 276:19836–19844.
20. Betermier, M., I. Poquet, R. Alazard, and M. Chandler. 1993. Involvement of *Escherichia coli* FIS protein in maintenance of bacteriophage mu lysogeny by the repressor: control of early transcription and inhibition of transposition. J. Bacteriol. 175:3798–3811.
21. Betermier, M., P. Rousseau, R. Alazard, and M. Chandler. 1995. Mutual stabilisation of bacteriophage Mu repressor and histone-like proteins in a nucleoprotein structure. J. Mol. Biol. 249:332–341.
22. Bidnenko, E. M., V. Z. Akhverdian, and V. N. Krylov. 2000. Transcriptional mapping and study of transcription regulation of the *Pseudomonas aeruginosa* phage-transposon D3112. Genetika 36:1645–1655.
23. Bölker, M., F. G. Wulczyn, and R. Kahmann. 1989. Role of bacteriophage Mu C protein in activation of the *mom* gene promoter. J. Bacteriol. 171:2019–2027.
24. Bukhari, A. I. 1975. Reversal of mutator phage Mu integration. J. Mol. Biol. 96:87–99.
25. Bukhari, A. I., and A. L. Taylor. 1975. Influence of insertions on packaging of host sequences covalently linked to bacteriophage Mu DNA. Proc. Natl. Acad. Sci. USA 72:4399–4403.
26. Bukhari, A. I., and D. Zipser. 1972. Random insertion of Mu-1 DNA within a single gene. Nat. New Biol. 236:240–243.
27. Bukhari, A. I., S. Froshauer, and M. Botchan. 1976. Ends of bacteriophage mu DNA. Nature 264:580–583.
28. Burns, C. M., H. L. Chan, and M. S. DuBow. 1990. In vitro maturation and encapsidation of the DNA of transposable Mu-like phage D108. Proc. Natl. Acad. Sci. USA 87:6092–6096.
29. Campbell, A., and D. Botstein. 1983. Bacteriophage 1, pp. 65–103. *In* J. A. Shapiro (ed.) Mobile Genetic Elements. Academic Press, New York.

30. Chaconas, G. 1987. Transposition of Mu DNA in vivo, pp. 137–157. *In* N. Symonds, A. Toussaint., P. van de Putte, and M. H. Howe (eds.) Phage Mu. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
31. Chaconas, G., and R. Harshey. 2002. Transposition of phage Mu DNA, pp. 384–402. *In*. N. L. Craig et al. eds. ASM Press, Washington, D.C.
32. Chaconas, G., B. D. Lavoie, and M. A. Watson. 1996. DNA transposition: jumping gene machine, some assembly required. Curr. Biol. 6:817–820.
33. Chaconas, G., D. L. Kennedy, and D. Evans. 1983. Predominant integration end products of infecting bacteriophage Mu DNA are simple insertions with no preference for integration of either Mu DNA strand. Virology. 128:48–59.
34. Chaconas, G., E. B. Giddens, J. L. Miller, and G. Gloor. 1985. A truncated form of the bacteriophage Mu B protein promotes conservative integration, but not replicative transposition, of Mu DNA. Cell 41:857–865.
35. Chaconas, G., F. J. de Bruijn, M. J. Casadaban, J. R. Lupski, T. J. Kwoh, R. M. Harshey, M. S. DuBow, and A. I. Bukhari. 1981. In vitro and in vivo manipulations of bacteriophage Mu DNA: cloning of Mu ends and construction of mini-Mu's carrying selectable markers. Gene 13:37–46.
36. Chaconas, G., G. Gloor, and J. L. Miller. 1985. Amplification and purification of the bacteriophage Mu encoded B transposition protein. J. Biol. Chem. 260:2662–2669.
37. Chaconas, G., G. Gloor, J. L. Miller, D. L. Kennedy, E. B. Giddens, and C. R. Nagainis. 1984. Transposition of bacteriophage mu DNA: expression of the A and B proteins from lambda pL and analysis of infecting mu DNA. Cold Spring Harb. Symp. Quant. Biol. 49:279–284.
38. Chase, C. D., and R. H. Benzinger. 1982. Transfection of *Escherichia coli* spheroplasts with a bacteriophage Mu DNA–protein complex. J. Virol. 42:176–185.
39. Coelho, A., S. Maynard-Smith, and N. Symonds. 1982. Abnormal cointegrate structures mediated by gene B mutants of phage Mu: their implications with regard to gene function. Mol. Gen. Genet. 185:356–662.
40. Coros, C. J., Y. Sekino, T. A. Baker, and G. Chaconas. 2003. Effect of mutations in the C-terminal domain of Mu B on DNA binding and interactions with Mu A transposase. J. Biol. Chem. 278:31210–31217.
41. Couturier, M., and F. Van Vliet. 1974. Vegetative recombination in bacteriophage Mu-1. Virology 60:1–8.
42. Craig, N. L. 1995. Unity in transposition reactions. Science 270:253–254.
43. Craigie, R., and K. Mizuuchi. 1985. Cloning of the A gene of bacteriophage Mu and purification of its product, the Mu transposase. J. Biol. Chem. 260:1832–1835.
44. Craigie, R., and K. Mizuuchi. 1985. Mechanism of transposition of bacteriophage Mu: structure of a transposition intermediate. Cell 41:867–876.
45. Craigie, R., and K. Mizuuchi. 1987. Transposition of Mu DNA: joining of Mu to target DNA can be uncoupled from cleavage at the ends of Mu. Cell 51:493–501.
46. Craigie, R., M. Mizuuchi, and K. Mizuuchi. 1984. Site-specific recognition of the bacteriophage Mu ends by the Mu A protein. Cell 39:387–394.
47. Daniell, E., and J. Abelson. 1973. Lac messenger RNA in *lac* Z gene mutants of *Escherichia coli* caused by insertion of bacteriophage Mu. J. Mol. Biol. 76:319–322.
48. Daniell, E., D. E. Kohne, and J. Abelson. 1975. Characterization of the inhomogeneous DNA in virions of bacteriophage Mu by DNA reannealing kinetics. J. Virol. 15:739–743.
49. Daniell, E., J. Abelson, J. S. Kim, and N. Davidson. 1973. Heteroduplex structures of bacteriophage Mu DNA. Virology 51:237–239.
50. Daniell, E., R. Roberts, and J. Abelson. 1972. Mutations in the lactose operon caused by bacteriophage Mu. J. Mol. Biol. 69:1–8.
51. Daniell, E., W. Boram, and J. Abelson. 1973. Genetic mapping of the inversion loop in bacteriophage Mu DNA. Proc. Natl. Acad. Sci. USA 70:2153–2156.
52. Dhillon, T. S., and E. K. Dhillon. 1972. Studies on bacteriophage distribution. II. Isolation and host range based classification of phages active on three species of Enterobacteriaceae. Jpn. J. Microbiol. 16:297–306.
53. Ding, Z. M., R. M. Harshey, and L. H. Hurley. 1993. (+)-CC-1065 as a structural probe of Mu transposase-induced bending of DNA: overcoming limitations of hydroxyl–radical footprinting. Nucleic Acids Res. 21:4281–4287.
54. Drlica, K. 1984. Biology of bacterial deoxyribonucleic acid topoisomerases. Microbiol. Rev. 48:273–289.
55. Egner, C., and D. E. Berg. 1981. Excision of transposon Tn5 is dependent on the inverted repeats but not on the transposase function of Tn5. Proc. Natl. Acad. Sci. USA 78:459–463.
56. Fabozzi, G., L. Paolozzi, and P. Ghelardini. 1998. Regulation of the bacteriophage Mu *gem* operon. Virology 241:73–79.
57. Faelen, M., A. Toussain, M. van Montagu, S. van den Elsacker, G. Engler, and J. Schell. 1977. In vivo genetic engineering: the Mu-mediated transposition of chromosomal DNA segments onto transmissible plasmids, pp. 521–530. *In* A. I. Bukhari et al. (eds.) DNA Insertion Elements, Plasmids and Episomes. Cold Spring Harbor Laboratory press , Cold Spring Harbor, N.Y.
58. Faelen, M., and A. Toussaint. 1973. Isolation of conditional defective mutants of temperate phage Mu-1 and deletion mapping of the Mu-1 prophage. Virology 54:117–124.
59. Faelen, M., and A. Toussaint. 1976. Bacteriophage Mu-1: a tool to transpose and to localize bacterial genes. J. Mol. Biol. 104:525–539.
60. Faelen, M., and A. Toussaint. 1978. Stimulation of deletions in the *Escherichia coli* chromosome by partially induced Mu*c*ts62 prophages. J. Bacteriol. 136:477–483.
61. Faelen, M., and A. Toussaint. 1980. Inversion induced by temperature bacteriophage mu-1 in the chromosome of *Escherichia coli* K-12. J. Bacteriol. 142:391–399.
62. Falconi, M., V. McGovern, C. Gualerzi, D. Hillyard, and N. P. Higgins. 1991. Mutations altering chromosomal protein H-NS induce mini-Mu transposition. New Biol. 3:615–625.
63. Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, J. M. Merrick, et al. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269:496–512.

64. Gama, M. J., A. Toussaint, and N. P. Higgins. 1992. Stabilization of bacteriophage Mu repressor-operator complexes by the *Escherichia coli* integration host factor protein. Mol. Microbiol. 6:1715–1722.
65. Gawron-Burke, C., and D. B. Clewell. 1982. A transposon in *Streptococcus faecalis* with fertility properties. Nature 300:281–284.
66. Gerdes, J. C., and W. R. Romig. 1975. Genetic basis of toxin production and pathogenesis in *Vibrio cholerae*: evidence against phage conversion. Infect. Immun. 11:445–452.
67. Geuskens, V., A. Mhammedi-Alaoui, L. Desmet, and A. Toussaint. 1992. Virulence in bacteriophage Mu: a case of *trans*-dominant proteolysis by the *Escherichia coli* Clp serine protease. EMBO J. 11:5121–5127.
68. Geuskens, V., J. L. Vogel, R. Grimaud, L. Desmet, N. P. Higgins, and A. Toussaint. 1991. Frameshift mutations in the bacteriophage Mu repressor gene can confer a *trans*-dominant virulent phenotype to the phage. J. Bacteriol. 173:6578–6585.
69. Ghelardini, P., J. C. Liebart, G. Di Zenzo, G. Micheli, R. D'Ari, and L. Paolozzi. 1994. A novel illegitimate recombination event: precise excision and reintegration with the Mu *gem* mutant prophage. Mol. Microbiol. 13:709–718.
70. Ghelardini, P., J. C. Liebart, L. Paolozzi, and A. M. Pedrini. 1989. Suppression of the thermosensitive DNA ligase mutations in *Escherichia coli* K12 through modulation of gene expression induced by phage Mu. Mol. Gen. Genet. 216:31–36.
71. Gindlesperger, T. L., and S. Hattman. 1994. In vitro transcriptional activation of the phage Mu mom promoter by C protein. J. Bacteriol. 176:2885–2891.
72. Giphart-Gassler, M., and P. Van de Putte. 1979. Thermoinducible expression of cloned early genes of bacteriophage Mu. Gene 7:33–50.
73. Giphart-Gassler, M., J. Reeve, and P. van de Putte. 1981. Polypeptides encoded by the early region of bacteriophage Mu synthesized in minicells of *Escherichia coli*. J. Mol. Biol. 145:165–191.
74. Giusti, M., G. Di Lallo, P. Ghelardini, and L. Paolozzi. 1990. The bacteriophage Mu gem gene: a positive regulator of the C operon required for normal levels of late transcription. Virology 179:694–700.
75. Gloor, G., and G. Chaconas. 1986. The bacteriophage Mu N gene encodes the 64-kDa virion protein which is injected with, and circularizes, infecting Mu DNA. J. Biol. Chem. 261:16682–16688.
76. Gomez-Gomez, J. M., J. Blazquez, F. Baquero, and J. L. Martinez. 1997. H-NS and RpoS regulate emergence of Lac Ara+ mutants of *Escherichia coli* MCS2. J. Bacteriol. 179:4620–4622.
77. Goosen, N., and P. van de Putte. 1984. Hek: an *Escherichia coli* function involved in functional expression of the *kil* gene of bacteriophage Mu. Mol. Gen. Genet. 196:170–172.
78. Goosen, N., and P. van de Putte. 1984. Regulation of Mu transposition. I. Localization of the presumed recognition sites for HimD and Ner functions controlling bacteriophage Mu transcription. Gene 30:41–46.
79. Goosen, N., and P. van de Putte. 1986. Role of ner protein in bacteriophage Mu transposition. J. Bacteriol. 167:503–507.
80. Goosen, N., and P. van de Putte. 1987. Regulation of transcription, pp. 41–52. *In* N. Symonds, A. Toussaint., P. van de Putte, and M. H. Howe (eds.) phage Mu. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
81. Goosen, N., M. van Heuvel, G. F. Moolenaar, and P. van de Putte. 1984. Regulation of Mu transposition. II. The Escherichia coli HimD protein positively controls two repressor promoters and the early promoter of bacteriophage Mu. Gene 32:419–426.
82. Gottesman, S., E. Roche, Y. Zhou, and R. T. Sauer. 1998. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. Genes Dev. 12:1338–1347.
83. Greene, E. C., and Mizuuchi, K. 2002. Direct observation of single MuB polymers: evidence for a DNA-dependent conformational change for generating an active target complex. Mol Cell 9:1079–1089.
84. Greene, E. C., and Mizuuchi, K. 2002. Dynamics of a protein polymer: the assembly and disassembly pathways of the MuB transposition target complex. EMBO J. 21:1477–86.
85. Greene, E. C., and Mizuuchi, K. 2004. Visualizing the assembly and disassembly mechanisms of the MuB transposition targeting complex. J. Biol. Chem. 279: 16736–16743.
86. Grimaud, R., M. Kessel, F. Beuron, A. Steven, and M. Maurizi. 1998. Enzymatic and structural similarities between the *Escherichia coli* ATP-dependent proteases, ClpXP and ClpAP. J. Biol. Chem. 273:12476–12481.
87. Groenen, M. A., and P. van de Putte. 1985. Mapping of a site for packaging of bacteriophage Mu DNA.Virology 144:520–522.
88. Grundy, F. J., and M. M. Howe. 1984. Involvement of the invertible G segment in bacteriophage mu tail fiber biosynthesis. Virology 134:296–317.
89. Harel, J., L. Duplessis, J. S. Kahn, and M. S. DuBow. 1990. The *cis*-acting DNA sequences required in vivo for bacteriophage Mu helper-mediated transposition and packaging. Arch. Microbiol. 154:67–72.
90. Harshey, R. M. 1984. Transposition without duplication of infecting bacteriophage Mu DNA. Nature 311:580–581.
91. Harshey, R. M. 1987. Integration of infecting Mu DNA, pp. 111–135. *In* N. Symonds, A. Toussaint., P. van de Putte, and M. H. Howe (eds.) Phage Mu. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
92. Harshey, R. M. 1988. Phage Mu, pp. 193–234. *In* R. Calendar (ed.) The Bacteriophages. Plenum Press, New York.
93. Harshey, R. M., and A. I. Bukhari. 1983. Infecting bacteriophage mu DNA forms a circular DNA–protein complex. J. Mol. Biol. 167:427–441.
94. Harshey, R. M., E. D. Getzoff, D. L. Baldwin, J. L. Miller, and G. Chaconas. 1985. Primary structure of phage mu transposase: homology to mu repressor. Proc. Natl. Acad. Sci. USA 82:7676–7680.
95. Hattman, S. 1979. Unusual modification of bacteriophage Mu DNA. J. Virol. 32:468–475.
96. Hattman, S. 1980. Specificity of the bacteriophage Mu mom$^+$ -controlled DNA modification. J. Virol. 34:277–279.

97. Hattman, S. 1999. Unusual transcriptional and translational regulation of the bacteriophage Mu mom operon. Pharmacol. Ther. 84:367–88.

98. Hattman, S., and J. Ives. 1984. S1 nuclease mapping of the phage Mu mom gene promoter: a model for the regulation of mom expression. Gene 29:185–198.

99. Hattman, S., J. Ives, W. Margolin, and M. M. Howe. 1985. Regulation and expression of the bacteriophage Mu *mom* gene: mapping of the transactivation (Dad) function to the *C* region. Gene 39:71–76.

100. Hattman, S., J. Ives, L. Wall, and S. Maric. 1987. The bacteriophage Mu com gene appears to specify a translation factor required for mom gene expression. Gene 55:345–351.

101. Hattman, S., L. Newman, H. M. Murthy, and V. Nagaraja. 1991. Com, the phage Mu mom translational activator, is a zinc-binding protein that binds specifically to its cognate mRNA. Proc. Natl. Acad. Sci. USA 88:10027–10031.

102. Hattman, S., M. Goradia, C. Monaghan, and A. I. Bukhari. 1983. Regulation of the DNA-modification function of bacteriophage Mu. Cold Spring Harb. Symp. Quant. Biol. 47:647–653.

103. Hayashi, T., K. Makino, M. Ohnishi, K. Kurokawa, K. Ishii, K. Yokoyama, et al. 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K12. DNA Res. 8:11–22.

104. Heisig, P., and R. Kahmann. 1986. The sequence and mom-transactivation function of the C gene of bacteriophage Mu. Gene 43:59–67.

105. Hendrix, R. W., J. G. Lawrence, G. F. Hatfull, and S. Casjens. 2000. The origins and ongoing evolution of viruses. Trends Microbiol. 8:504–508.

106. Howe, M. M. 1973. Prophage deletion mapping of bacteriophage Mu-1. Virology 54:93–101.

107. Howe, M. M. 1973. Transduction by bacteriophage MU-1. Virology 55:103–117.

108. Howe, M. M., and E. G. Bade. 1975. Molecular biology of bacteriophage Mu. Science 190:624–632.

109. Hsu, M. T., and N. Davidson. 1972. Structure of inserted bacteriophage Mu-1 DNA and physical mapping of bacterial genes by Mu-1 DNA insertion. Proc. Natl. Acad. Sci. USA 69:2823–2827.

110. Hsu, M. T., and N. Davidson. 1974. Electron microscope heteroduplex study of the heterogeneity of Mu phage and prophage DNA. Virology 58:229–239.

111. Hung, L. H., G. Chaconas, and G. S. Shaw. 2000. The solution structure of the C-terminal domain of the Mu B transposition protein. EMBO J. 19:5625–5634.

112. Jacob, F., and E. Wollman. 1954. Spontaneous induction of the development of bacteriophage lambda during genetic recombination in *Escherichia coli* K12. C. R. Hebd. Seances. Acad. Sci. 239:317–319.

113. Jiang, H. J. Y. Yanag, and R. M. Harshey. 1999. Crisscrossed interactions between the enhancer and the *att* sites of phage Mu during DNA transposition. EMBO J. 18: 3845–3855.

114. Johnson, S. R., and W. R. Romig. 1981. *Vibrio cholerae* conjugative plasmid pSJ15 contains transposable prophage dVcA1. J. Bacteriol. 146:632–638.

115. Jones, J. M., and H. Nakai. 2000. PriA and phage T4 gp59: factors that promote DNA replication on forked DNA substrates microreview. Mol. Microbiol. 36:519–527.

116. Jones, J. M., D. J. Welty, and H. Nakai. 1998. Versatile action of *Escherichia coli* ClpXP as protease or molecular chaperone for bacteriophage Mu transposition. J. Biol. Chem. 273:459–465.

117. Jordan, E., H. Saedler, and P. Starlinger. 1968. O° and strong-polar mutations in the *gal* operon are insertions. Mol. Gen. Genet. 102:353–363.

118. Juhala, R. J., M. E. Ford, R. L. Duda, A. Youlton, G. F. Hatfull, and R. W. Hendrix. 2000. Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. J. Mol. Biol. 299:27–51.

119. Kahmann, R. 1984. The *mom* gene of bacteriophage Mu. Curr. Top. Microbiol. Immunol. 108:29–47.

120. Kahmann, R., and D. Kamp 1979. Nucleotide sequences of the attachment sites of bacteriophage Mu DNA. Nature 280:247–250.

121. Kahmann, R., and S. Hattman. 1987. Regulation and expression of the *mom* gene, pp. 93–109. *In* N. Symonds, A. Toussaint., P. van de Putte, and M. H. Howe (eds.) Phage Mu. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

122. Kahmann, R., D. Kamp, and D. Zipser. 1976. Transfection of *Escherichia coli* by Mu DNA. Mol. Gen. Genet. 149:323–328.

123. Kahmann, R., F. Rudt, and G. Mertens. 1984. Substrate and enzyme requirements for in vitro site-specific recombination in bacteriophage mu. Cold Spring Harb. Symp. Quant. Biol. 49:285–294.

124. Kahmann, R., F. Rudt, C. Koch, and G. Mertens. 1985. G. inversion in bacteriophage Mu DNA is stimulated by a site within the invertase gene and a host factor. Cell 41:771–780.

125. Kahmann, R., A. Seiler, F. G. Wulczyn, and E. Pfaff. 1985. The *mom* gene of bacteriophage Mu: a unique regulatory scheme to control a lethal function. Gene 39:61–70.

126. Kahmeyer-Gabbe, M., and M. M. Howe. 1996. Regulatory factors acting at the bacteriophage Mu middle promoter. J. Bacteriol. 178:1585–1592.

127. Kamp, D. 1981. Invertible DNA: the G segment of bacteriophage Mu, pp. 73–75. *In* D. Schlessinger (ed.) Microbiology. American Society for Microbiology, Washington, D.C.

128. Kamp, D. 1987. The evolution of Mu, pp. 259–269. *In* N. Symonds, A. Toussaint., P. van de Putte, and M. H. Howe (eds.) Phage Mu. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

129. Kamp, D., and R. Kahmann. 1981. The relationship of two invertible segments in bacteriophage Mu and *Salmonella typhimurium* DNA. Mol. Gen. Genet. 184:564–566.

130. Kamp, D., E. Kardas, W. Ritthaler, R. Sandulache, R. Schmucker, and B. Stern. 1984. Comparative analysis of invertible DNA in phage genomes. Cold Spring Harb. Symp. Quant. Biol. 49:301–311.

131. Kamp, D., L. T. Chow, T. R. Broker, D. Kwoh, D. Zipser, and R. Kahmann. 1979. Site-specific recombination in phage Mu. Cold Spring Harb. Symp. Quant. Biol. 43: 1159–1167.

132. Kamp, D., R. Kahmann, D. Zipser, T. R. Broker, and L. T. Chow. 1978. Inversion of the G DNA segment of phage Mu controls phage infectivity. Nature 271:577–580.

133. Kanaar, R., P. van de Putte, and N. R. Cozzarelli. 1988. Gin-mediated DNA inversion: product structure and the mechanism of strand exchange. Proc. Natl. Acad. Sci. USA 85:752–756.

134. Keiler, K. C., P. R. Waller, and R. T. Sauer. 1996. Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. Science 271:990–993.

135. Khatoon, H., G. Chaconas, M. DuBow, and A. I. Bukhari. 1979. The Mu paradox: excision versus replication. ICN-UCLA Symp. Mol. Cell Biol. 15:143–154.

136. Kim, K., S. Y. Namgoong, M. Jayaram, and R. M. Harshey. 1995. Step-arrest mutants of phage Mu transposase. Implications in DNA-protein assembly, Mu end cleavage, and strand transfer. J. Biol. Chem. 270:1472–1479.

137. Klee, S. R., X. Nassif, B. Kusecek, P. Merker, J. L. Beretti, M. Achtman, and C. R. Tinsley. 2000. Molecular and biological analysis of eight genetic islands that distinguish *Neisseria meningitidis* from the closely related pathogen *Neisseria gonorrhoeae*. Infect. Immun. 68: 2082–2095.

138. Koch, C., and R. Kahmann. 1986. Purification and properties of the *Escherichia coli* host factor required for inversion of the G segment in bacteriophage Mu. J. Biol. Chem. 261:15673–15678.

139. Koch, C., G. Mertens, F. Rudt, R. Khamann, R. Kanaar, R. Plasterk, R. Sandulache, and D. Kamp. 1987. The invertible G segment, pp. 75–91. *In* N. Symonds, A. Toussaint., P. van de Putte, and M.H. Howe (eds.) Phage Mu. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

140. Koch, C., O. Ninnemann, H. Fuss, and R. Kahmann. 1991. The N-terminal part of the *E. coli* DNA binding protein FIS is essential for stimulating site-specific DNA inversion but is not required for specific DNA binding. Nucleic Acids Res. 19:5915–5922.

141. Krause, H. M., M. R. Rothwell, and N. P. Higgins. 1983. The early promoter of bacteriophage Mu: definition of the site of transcript initiation. Nucleic Acids Res. 11:5483–5495.

142. Krause, H. M., and N. P. Higgins. 1986. Positive and negative regulation of the Mu operator by Mu repressor and *Escherichia coli* integration host factor. J. Biol. Chem. 261:3744–3752.

143. Krementsova, E., M. J. Giffin, D. Pincus, and T. A. Baker. 1998. Mutational analysis of the Mu transposase. Contributions of two distinct regions of domain II to recombination. J. Biol. Chem. 273:31358–31365.

144. Kruklitis, R., and H. Nakai. 1994. Participation of the bacteriophage Mu A protein and host factors in the initiation of Mu DNA synthesis in vitro. J. Biol. Chem. 269:16469–16477.

145. Kruklitis, R., D. J. Welty, and H. Nakai. 1996. ClpX protein of *Escherichia coli* activates bacteriophage Mu transposase in the strand transfer complex for initiation of Mu DNA synthesis. EMBO J. 15:935–944.

146. Krylov, V. N., V. G. Bogush, A. S. Ianenko, and N. B. Kirsanov. 1980. *Pseudomonas aeruginosa* bacteriophages with DNA structure similar to the DNA structure of Mu1 phage. II. Evidence for similarity between D3112, B3, and B39 bacteriophages: analysis of DNA splits by restriction endonucleases, isolation of D3112 and B3 recombinant phages. Genetika 16:975–984.

147. Kwoh, D., and D. Zipser. 1979. Specific binding of mu repressor to DNA. Nature 277:489–491.

148. Laachouch, J. E., L. Desmet, V. Geuskens, R. Grimaud, and A. Toussaint. 1996. Bacteriophage Mu repressor as a target for the *Escherichia coli* ATP-dependent Clp protease. EMBO J. 15:437–444.

149. Lamrani, S., C. Ranquet, M. J. Gama, H. Nakai, J. A. Shapiro, A. Toussaint, and G. Maenhaut–Michel. 1999. Starvation-induced Mucts62-mediated coding sequence fusion: a role for ClpXP, Lon, RpoS and Crp. Mol. Microbiol. 32:327–343.

150. Lavoie, B. D., B. S. Chan, R. G. Allison, and G. Chaconas. 1991. Structural aspects of a higher order nucleoprotein complex: induction of an altered DNA structure at the Mu–host junction of the Mu type 1 transpososome. EMBO J. 10:3051–3059.

151. Lavoie, B. D., and G. Chaconas. 1990. Immunoelectron microscopic analysis of the A, B, and HU protein content of bacteriophage Mu transpososomes. J. Biol. Chem. 265:1623–1627.

152. Lavoie, B. D., and G. Chaconas. 1995. Transposition of phage Mu DNA. Curr. Top. Microbiol. Immunol. 204:83–99.

153. Lavoie, B.D., and G. Chaconas. 1996. Transposition of phage Mu DNA. Curr. Top. Microbiol. Immunol. 204:83–102.

154. Leach, D., and N. Symonds. 1979. The isolation and characterisation of a plaque-forming derivative of bacteriophage Mu carrying a fragment of Tn3 conferring ampicillin resistance. Mol. Gen. Genet. 172:179–184.

155. Leung, P. C., and R. M. Harshey. 1991. Two mutations of phage mu transposase that affect strand transfer or interactions with B protein lie in distinct polypeptide domains. J. Mol. Biol. 219:189–199.

156. Leung, P. C., D. B. Teplow, and R. M. Harshey. 1989. Interaction of distinct domains in Mu transposase with Mu DNA ends and an internal transpositional enhancer. Nature 338:656–658.

157. Levchenko, I., L. Luo, and T. A. Baker. 1995. Disassembly of the Mu transposase tetramer by the ClpX chaperone. Genes Dev. 9:2399–2408.

158. Liebart, J. C., P. Ghelardini, and L. Paolozzi. 1982. Conservative integration of bacteriophage Mu DNA into pBR322 plasmid. Proc. Natl. Acad. Sci. USA 79: 4362–4366.

159. Ljungquist, E., and A. I. Bukhari. 1977. State of prophage Mu DNA upon induction. Proc. Natl. Acad. Sci. USA 74:3143–3147.

160. Ljungquist, E., and A. I. Bukhari. 1979. Behavior of bacteriophage Mu DNA upon infection of *Escherichia coli* cells. J. Mol. Biol. 133:339–357.

161. Ljungquist, E., H. Khatoon, M. DuBow, L. Ambrosio, F. De Bruijn, and A. I. Bukhari. 1979. Integration of bacteriophage mu DNA. Cold Spring Harb. Symp. Quant. Biol. 43:1151–1158.

162. Lovell, S., I. Y. Goryshin, W. R. Reznikoff, and I. Rayment. 2002. Two-metal active site binding of a Tn5 transposase synaptic complex. Nat. Struct. Biol. 9:278–281.
163. Lyon, B. R., and R. Skurray. 1987. Antimicrobial resistance of *Staphylococcus aureus*: genetic basis. Microbiol. Rev. 51:88–134.
164. Manna, D., and N. P. Higgins. 1999. Phage Mu transposition immunity reflects supercoil domain structure of the chromosome. Mol. Microbiol. 32:595–606.
165. Margolin, W., and M. M. Howe. 1986. Localization and DNA sequence analysis of the C gene of bacteriophage Mu, the positive regulator of Mu late transcription. Nucleic Acids Res. 14:4881–4897.
166. Margolin, W., and M. M. Howe. 1990. Activation of the bacteriophage Mu lys promoter by Mu C protein requires the sigma 70 subunit of *Escherichia coli* RNA polymerase. J. Bacteriol. 172:1424–1429.
167. Margolin, W., G. Rao, and M. M. Howe. 1989. Bacteriophage Mu late promoters: four late transcripts initiate near a conserved sequence. J. Bacteriol. 171: 2003–2018.
168. Marrs, C. F., and M. M. Howe. 1983. *Ava*II and *Bgl*I restriction maps of bacteriophage Mu. Virology 126: 563–575.
169. Marrs, C. F., and M. M. Howe. 1990. Kinetics and regulation of transcription of bacteriophage Mu. Virology. 174:192–203.
170. Martuscelli, J., A. L. Taylor, D. J. Cummings, V. A. Chapman, S. S. DeLong, and L. Canedo. 1971. Electron microscopic evidence for linear insertion of bacteriophage Mu-1 in lysogenic bacteria. J. Virol. 8:551–563.
171. Masignani, V., M. M. Giuliani, H. Tettelin, M. Comanducci, R. Rappuoli, and V. Scarlato. 2001. Mu-like Prophage in serogroup B *Neisseria meningitidis* coding for surface-exposed antigens. Infect. Immun. 69:2580–2588.
172. Mathee, K., and M. M. Howe. 1990. Identification of a positive regulator of the Mu middle operon. J. Bacteriol. 172:6641–6650.
173. Mathee, K., and M. M. Howe. 1993. Bacteriophage Mu Mor protein requires sigma 70 to activate the Mu middle promoter. J. Bacteriol. 175:5314–5323.
174. Mathee, K., and M. M. Howe. 1993. The bacteriophage Mu middle operon: essential and nonessential functions. Virology 196:712–721.
175. Maxwell, A., R. Craigie, and K. Mizuuchi. 1987. B protein of bacteriophage Mu is an ATPase that preferentially stimulates intermolecular DNA strand transfer. Proc. Natl. Acad. Sci. USA 84:699–703.
176. McClintock, B. 1956. Controlling elements and the gene. Cold Spring Harb. Symp. Quant. Biol. 21:197–216.
177. Mhammedi-Alaoui, A., M. Pato, M. J. Gama, and A. Toussaint. 1994. A new component of bacteriophage Mu replicative transposition machinery: the *Escherichia coli* ClpX protein. Mol. Microbiol. 11:1109–1116.
178. Mise, K. 1971. Isolation and characterization of a new generalized transducing bacteriophage different from P1 in *Escherichia coli*. J. Virol. 7:168–175.
179. Mizuuchi, K. 1983. In vitro transposition of bacteriophage Mu: a biochemical approach to a novel replication reaction. Cell 35:785–794.
180. Mizuuchi, K. 1992. Transpositional recombination: mechanistic insights from studies of mu and other elements. Annu. Rev. Biochem. 61:1011–1051.
181. Mizuuchi, M., and K. Mizuuchi. 1989. Efficient Mu transposition requires interaction of transposase with a DNA sequence at the Mu operator: implications for regulation. Cell 58:399–408.
182. Mizuuchi, M., and K. Mizuuchi. 2001. Conformational isomerization in phage Mu transpososome assembly: effects of the transpositional enhancer and of MuB. EMBO J. 20:6927–6935.
183. Mizuuchi, M., T. A. Baker, and K. Mizuuchi. 1992. Assembly of the active form of the transposase–Mu DNA complex: a critical control point in Mu transposition. Cell 70:303–311.
184. Mizuuchi, M., T. A. Baker, and K. Mizuuchi. 1995. Assembly of phage Mu transpososomes: cooperative transitions assisted by protein and DNA scaffolds. Cell 83:375–385.
185. Morgan, G. J., G. F. Hatfull, S. Casjens, and R. W. Hendrix. 2002. Bacteriophage Mu genome sequence: analysis and comparison with Mu-like prophages in *Haemophilus*, *Neisseria* and *Deinococcus*. J. Mol. Biol. 317:337–359.
186. Naigamwalla, D. Z., and G. Chaconas. 1997. A new set of Mu DNA transposition intermediates: alternate pathways of target capture preceding strand transfer. EMBO J. 16:5227–5234.
187. Nakai, H., and R. Kruklitis. 1995. Disassembly of the bacteriophage Mu transposase for the initiation of Mu DNA replication. J. Biol. Chem. 270:19591–19598.
188. Nakai, H., V. Doseeva, and J. M. Jones. 2001. Handoff from recombinase to replisome: insights from transposition. Proc. Natl. Acad. Sci. USA 98:8247–8254.
189. Nakayama, C., D. B. Teplow, and R. M. Harshey. 1987. Structural domains in phage Mu transposase: identification of the site-specific DNA-binding domain. Proc. Natl. Acad. Sci. USA 84:1809–1813.
190. Neuwald, A. F., L. Aravind, J. L. Spouge, and E. V. Koonin. 1999. AAA+: a class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. Genome Res. 9:27–43.
191. O' Day, K. J., D. W. Schultz, and M. M. Howe. 1978. Search for integration-deficient mutants of bacteriophage Mu, pp. 48–51. *In* D. Schlessinger (ed.) Microbiology. American Society for Microbiology, Washington, D.C.
192. O'Day, K., D. Schultz, W. Ericsen, L. Rawluk, and M. Howe. 1979. Correction and refinement of the genetic map of bacteriophage Mu. Virology 93:320–328.
193. O'Handley, D., and H. Nakai. 2002. Derepression of bacteriophage mu transposition functions by truncated forms of the immunity repressor. J. Mol. Biol. 322:311–324.
194. P. van Ulsen, M. Hillebrand, L. Zulianello, P. van de Putte, and N. Goosen. 1996. Integration host factor alleviates the H-NS-mediated repression of the early promoter of bacteriophage Mu. Mol. Microbiol. 21:567–578.
195. Paolozzi, L., and N. Symonds. 1987. The SER region, pp. 259–269. *In* N. Symonds, A. Toussaint., P. van de Putte, and M. H. Howe (eds.) Phage Mu. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
196. Paolozzi, L., and P. Ghelardini. 1992. A case of lysogenic conversion: modification of cell phenotype by constitutive

expression of the Mu *gem* operon. Res. Microbiol. 143:237–243.

197. Paolozzi, L., G. Fabozzi, and P. Ghelardini. 2000. Mu DNA reintegration upon excision: evidence for a possible involvement of nucleoid folding. Microbiology 146:591–598.
198. Parker, *cited in* S. R. Johnson and W. R. Romig. 1981. *Vibrio cholerae* conjugative plasmid pSJ15 contains transposable prophage dVcA1. J. Bacteriol. 146:632–638.
199. Parkhill, J. 2000. In defense of complete genomes. Nat. Biotechnol. 18:493–494.
200. Parkhill, J., M. Achtman, K. D. James, S. D. Bentley, C. Churcher, S. R. Klee, G. Morelli, D. Basham, D. Brown, T. Chillingworth, R. M. Davies, P. Davis, K. Devlin, T. Feltwell, N. Hamlin, S. Holroyd, K. Jagels, S. Leather, S. Moule, K. Mungall, M. A. Quail, M. A. Rajandream, K. M. Rutherford, M. Simmonds, J. Skelton, S. Whitehead, B. G. Spratt, and B. G. Barrell. 2000. Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. Nature 404:502–506.
201. Pathania, S., M. Jayaram, and R. M. Harshey. 2003. A unique right end-enhancer complex precedes synapsis of Mu ends: the enhancer is sequestered within the transpososome throughout transposition. EMBO J. 22:3725–3736.
202. Pato, M. L., and C. Reich. 1984. Stoichiometric use of the transposase of bacteriophage Mu. Cell 36:197–202.
203. Pato, M. L., and M. Banerjee. 1996. The Mu strong gyrase-binding site promotes efficient synapsis of the prophage termini. Mol. Microbiol. 22:283–292.
204. Pato, M. L., Howe, M. M., and N. P. Higgins, 1990. A DNAS-gyrase binding site at the center of the bacteriophage Mu genome is required for efficient replicative transposition. Proc. Natl. Acad. Sci. USA 87:8716–8620.
205. Plasterk, R. H., R. Kanaar, and P. van de Putte. 1984. A genetic switch in vitro: DNA inversion by Gin protein of phage Mu. Proc. Natl. Acad. Sci. USA 81:2689–2692.
206. Plasterk, R. H., T. A. Ilmer, and P. Van de Putte. 1983. Site-specific recombination by Gin of bacteriophage Mu: inversions and deletions. Virology 127:24–36.
207. Plasterk, R. H. A., H. Vollering, and P. van de Putte. 1984. Analysis of the methylation-regulated Mu *mom* transcript. Cell 36:189–196.
208. Priess, H., D. Kamp, R. Kahmann, B. Brauer, and H. Delius. 1982. Nucleotide sequence of the immunity region of bacteriophage Mu. Mol. Gen. Genet. 186:315–321.
209. Puspurs, A. H., N. J. Trun, and J. N. Reeve. 1983. Bacteriophage Mu DNA circularizes following infection of *Escherichia coli*. EMBO J. 2:345–352.
210. Raibaud, O., J. M. Clement, and M. Hofnung. 1979. Structure of the malB region in *Escherichia coli* K12. III. Correlation of the genetic map with the restriction map. Mol. Gen. Genet. 174:261–267.
211. Ramesh, V., and V. Nagaraja. 1996. Sequence-specific DNA binding of the phage Mu C protein: footprinting analysis reveals altered DNA conformation upon protein binding. J. Mol. Biol. 260:22–33.
212. Ranquet, C., J. Geiselmann, and A. Toussaint. 2001. The tRNA function of SsrA contributes to controlling repression of bacteriophage Mu prophage. Proc. Natl. Acad. Sci. USA 98:10220–10225.
213. Rehmat, S., and Shapiro, J. A. 1983. Insertion and replication of the *Pseudomonas aeruginosa* mutator phage D3112. Mol. Gen. Genet. 192:416–423.
214. Rice, P., and K. Mizuuchi. 1995. Structure of the bacteriophage Mu transposase core: a common structural motif for DNA transposition and retroviral integration. Cell 82:209–220.
215. Rice, P., R. Craigie, and D. R. Davies. 1996. Retroviral integrases and their cousins. Curr. Opin. Struct. Biol. 6:76–83.
216. Roldan, L. A., and T. A. Baker. 2001. Differential role of the Mu B protein in phage Mu integration vs. replication: mechanistic insights into two transposition pathways. Mol. Microbiol. 40:141–155.
217. Salmon, K. A., O. Freedman, B. W. Ritchings, and M. S. DuBow. 2000. Characterization of the lysogenic repressor (*c*) gene of the *Pseudomonas aeruginosa* transposable bacteriophage D3112. Virology 272:85–97.
218. Sandler, S. J., K. J. Marians, K. H. Zavitz, J. Coutu, M. A. Parent, and A. J. Clark. 1999. dnaC mutations suppress defects in DNA replication– and recombination-associated functions in priB and priC double mutants in *Escherichia coli* K-12. Mol. Microbiol. 34:91–101.
219. Schmucker, R., W. Ritthaler, B. Stern, and D. Kamp. 1986. DNA inversion in bacteriophage Mu: characterization of the inversion site. J. Gen. Virol. 67:1123–1133.
220. Shapiro, J. A. 1979. Molecular model for the transposition and replication of bacteriophage Mu and other transposable elements. Proc. Natl. Acad. Sci. USA 76:1933–1937.
221. Shapiro, J. A. 1993. A role for the Clp protease in activating Mu-mediated DNA rearrangements. J. Bacteriol. 175:2625–2631.
222. Shapiro, J. A., and D. Leach. 1990. Action of a transposable element in coding sequence fusions. Genetics 126: 293–299.
223. Shapiro, J. A., and N. P. Higgins. 1989. Differential activity of a transposable element in *Escherichia coli* colonies. J. Bacteriol. 171:5975–5986.
224. Schirmer, E., J. Glover, M. Singer, and S. Lindquist. 1996. HSP100/Clp proteins: a common mechanism explains diverse functions. Trends Biochem. Sci. 21:289–296.
225. Silhavy, T. J., E. Brickman, P. J. Bassford Jr., M. J. Casadaban, H. A. Schuman, V. Schwartz, L. Guarente, M. Schwartz, and J. R. Beckwith. 1979. Structure of the malB region in *E. coli* K12. Genetic map of the malE,F,G operon. Mol. Gen. Genet. 174:249–259.
226. Sokolsky, T. D., and T. A. Baker. 2003. DNA gyrase requirements distinguish the alternate pathways of Mu transposition. Mol. Microbiol. 47:397–409.
227. Stoddard, S. F., and M. M. Howe. 1987. DNA sequence within the Mu C operon. Nucleic Acids Res. 15:7198.
228. Stoddard, S. F., and M. M. Howe. 1990. Characterization of the C operon transcript of bacteriophage Mu. J. Bacteriol. 172:361–371.
229. Strzelecka, T. E., J. J. Hayes, G. M. Clore, and A. M. Gronenborn. 1995. DNA binding specificity of the Mu Ner protein. Biochemistry 34:2946–2955.
230. Sun, W., and S. Hattman. 1998. Bidirectional transcription in the mom promoter region of bacteriophage Mu. J. Mol. Biol. 284:885–892.

231. Surette, M. G., and G. Chaconas. 1992. The Mu transpositional enhancer can function in trans: requirement of the enhancer for synapsis but not strand cleavage. Cell 68:1101–1108.
232. Surette, M. G., S. J. Buch, and G. Chaconas. 1987. Transpososomes: stable protein–DNA complexes involved in the in vitro transposition of bacteriophage Mu DNA. Cell 49:253–262.
233. Swinton, D., S. Hattman, P. F. Crain, C. S. Cheng, D. L. Smith, and J. A. McCloskey. 1983. Purification and characterization of the unusual deoxynucleoside, α-*N*-(9-β-D-2′-deoxyribofuranosylpurin-6-yl) glycinamide, specified by the phage Mu modification function. Proc. Natl. Acad. Sci. USA 80:7400–7404.
234. Symonds, N., and A. Coelho. 1978. Role of the G segment in the growth of phage Mu. Nature 271:573–574.
235. Symonds, N., A. Toussaint, A., P. van de Putte, and W. V. Howes. 1987. Phage Mu. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
236. Taylor, A. L. 1963. Bacteriophage-induced mutation in *E. coli*. Proc. Natl. Acad. Sci. USA 50:1043–1051.
237. Teplow, D. B., C. Nakayama, P. C. Leung, and R. M. Harshey. 1988. Structure–function relationships in the transposition protein B of bacteriophage Mu. J. Biol. Chem. 263:10851–10857.
238. Tettelin, H., N. J. Saunders, J. Heidelberg, A. C. Jeffries, K. E. Nelson, J. A. Eisen, K. A. Ketchum, D. W. Hood, J. F. Peden, R. J. Dodson, W. C. Nelson, M. L. Gwinn, R. DeBoy, J. D. Peterson, E. K. Hickey, D. H. Haft, S. L. Salzberg, O. White, R. D. Fleischmann, B. A. Dougherty, T. Mason, A. Ciecko, D. S. Parksey, E. Blair, H. Cittone, E. B. Clark, M. D. Cotton, T. R. Utterback, H. Khouri, H. Qin, J. Vamathevan, J. Gill, V. Scarlato, V. Masignani, M. Pizza, G. Grandi, L. Sun, H. O. Smith, C. M. Fraser, E. R. Moxon, R. Rappuoli, and J. C. Venter. 2000. Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. Science 287:1809–1815.
239. To, C. M., A. Eisenstark, and H. Toreci. 1966. Structure of mutator phage Mu-1 of *Escherichia coli*. J. Ultrastruct. Res. 14:441–448.
240. Torti, F., C. Barksdale, and J. Abelson. 1970. Mu-1 bacteriophage DNA. Virology 41:567–568.
241. Toussaint, A., L. Desmet, F. van Gijsegem, and M. Faelen. 1981. Genetic analysis of mu or mini-mu containing F′ pro lac episomes after prophage induction. Mol. Gen. Genet. 181:201–206.
242. Toussaint, A., M. Faelen, L. Desmet, and B. Allet. 1983. The products of gene A of related phages, Mu and D108 differ in their specifities. Mol. Gen. Genet. 190:70–79.
243. Toussaint, A., and A. Résibois. 1983. Phage Mu: transposition as a life-style, pp. 105–158. *In* J. A. Shapiro Mobile Genetic Elements. Academic Press, New York.
244. Tu, G. F., G. E. Reid, J. G. Zhang, R. L. Moritz, and R. J. Simpson. 1995. C-terminal extension of truncated recombinant proteins in *Escherichia coli* with a 10Sa RNA decapeptide. J. Biol. Chem. 270:9322–9326.
245. van de Putte, P., M. Giphart-Gassler, N. Goosen, T. Goosen, and E. van Leerdam. 1981. Regulation of integration and replication functions of bacteriophage Mu. Cold Spring Harb. Symp. Quant. Biol. 45:347–353.
246. van de Putte, P., M. Giphart-Gassler, T. Goosen, A. van Meeteren, and C. Wijffelman. 1978. Is integration essential for Mu development?, pp. 33–40 *In* P. Hofschneider and P. Starlinger (eds.) Integration and Excision of DNA Molecules. Springer, Berlin.
247. van de Putte, P., S. Cramer, and M. Giphart-Gassler. 1980. Invertible DNA determines host specificity of bacteriophage mu. Nature 286:218–222.
248. van Gijsegem, F., A. Toussaint, and M. Casadaban. 1987. Mu as a genetic tool, pp. 215–250. *In* N. Symonds, A. Toussaint., P. van de Putte, and M. H. Howe (eds.) Phage Mu. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
249. van Leerdam, E., C. Karreman, and P. van de Putte. 1982. *Ner*, a *cro*-like function of bacteriophage Mu. Virology 123:19–28.
250. van Meeteren, A., and P. van de Putte. 1980. Transcription of bacteriophage Mu. I. Hybridization analysis of RNA made in vitro. Mol. Gen. Genet. 179:177–183.
251. van Rijn, P. A., N. Goosen, and P. van de Putte. 1988. Integration host factor of *Escherichia coli* regulates early- and repressor transcription of bacteriophage Mu by two different mechanisms. Nucleic Acids Res. 16:4595–4605.
252. van Rijn, P. A., N. Goosen, S. C. Turk, and P. van de Putte. 1989. Regulation of phage Mu repressor transcription by IHF depends on the level of the early transcription. Nucleic Acids Res. 17:10203–10212.
253. van Vliet, F., B. Silva, M. van Montagu, and J. Schell. 1978. Transfer of RP4::mu plasmids to *Agrobacterium tumefaciens*. Plasmid 1:446–455.
254. Vogel, J. L., Z. J. Li, M. M. Howe, A. Toussaint, and N. P. Higgins. 1991. Temperature-sensitive mutations in the bacteriophage Mu c repressor locate a 63–amino-acid DNA-binding domain. J. Bacteriol. 173:6568–6577.
255. Waggoner, B. T., and M. L. Pato. 1978. Early events in the replication of Mu prophage DNA. J. Virol. 27:587–594.
256. Waggoner, B., M. Pato, A. Toussaint, and M. Faelen. 1981. Replication of mini-Mu prophage DNA. Virology 113:379–387.
257. Wang, P. W., L. Chu, and D. S. Guttman. 2004. Complete sequence and evolutionary genomic analysis of the *Pseudomonas aeruginosa* transposabler bacteriophage D3112. J. Bacteriol. 186:400–410.
258. Wang, Z., S. Y. Namgoong, X. Zhang, and R. M. Harshey. 1996. Kinetic and structural probing of the precleavage synaptic complex (type 0) formed during phage Mu transposition. Action of metal ions and reagents specific to single-stranded DNA. J. Biol. Chem. 271:9619–9626.
259. Watson, M. A., and G. Chaconas. 1996. Three-site synapsis during Mu DNA transposition: a critical intermediate preceding engagement of the active site. Cell 85:435–445.
260. Welty, D. J., J. M. Jones, and H. Nakai. 1997. Communication of ClpXP protease hypersensitivity to bacteriophage Mu repressor isoforms. J. Mol. Biol. 272:31–41.
261. White, O., J. A. Eisen, J. F. Heidelberg, E. K. Hickey, J. D. Peterson, Dodson R. J., et al. 1999. Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1. Science 286:1571–1577.

262. Wijffelman, C. A., G. C. Westmaas, and P. van de Putte. 1972. Vegetative recombination of bacteriophage Mu-1 in *Escherichia coli*. Mol. Gen. Genet. 116:40–46.

263. Wijffelman, C. A., G. C. Westmaas, and P. van de Putte. 1973. Similarity of vegetative map and prophage map of bacteriophage Mu-1. Virology 54:125–134.

264. Wijffelman, C., and B. Lotterman. 1977. Kinetics of Mu DNA synthesis. Mol. Gen. Genet. 151:169–174.

265. Wijffelman, C., and P. van de Putte. 1974. Transcription of bacteriophage mu. An analysis of the transcription pattern in the early phase of phage development. Mol. Gen. Genet. 135:327–337.

266. Wijffelman, C., M. Gassler, W. F. Stevens, and P. van de Putte. 1974. On the control of transcription of bacteriophage Mu. Mol. Gen. Genet. 131:85–96.

267. Williams, T. L., and T. A. Baker. 2004. Reorganization of the Mu transpososome active sites during a cooperative transition between DNA cleavage and joining. J. Biol. Chem. 279:5135–5145.

268. Wlodawer, A. 1999. Crystal structures of catalytic core domains of retroviral integrases and role of divalent cations in enzymatic activity. Adv. Virus Res.52:335–350.

269. Woese, C. R. 1987. Bacterial evolution. Microbiol. Rev. 51:221–271.

270. Wu, Z., and G. Chaconas. 1994. Characterization of a region in phage Mu transposase that is involved in interaction with the Mu B protein. J. Biol. Chem. 269:28829–28833.

271. Wu, Z., and G. Chaconas. 1995. A novel DNA binding and nuclease activity in domain III of Mu transposase: evidence for a catalytic region involved in donor cleavage. EMBO J. 14:3835–3843.

272. Wulczyn, F. G., M. Bolker, and R. Kahmann. 1989. Translation of the bacteriophage Mu mom gene is positively regulated by the phage com gene product. Cell 57:1201–1210.

273. Wulczyn, F. G., and R. Kahmann. 1987. Post-transcriptional regulation of the bacteriophage Mu *mom* gene by the *com* gene product. Gene. 51:139–147.

274. Wulczyn, F. G., and R. Kahmann. 1991. Translational stimulation: RNA sequence and structure requirement for binding of Com protein. Cell 65:259–269.

275. Zipser, D., P. Moses, R. Kahmann, and D. Kamp. 1977. The molecular cloning of the immunity gene of phage Mu. Gene 2:263–271.

276. Zou, A. H., P. C. Leung, and R. M. Harshey. 1991. Transposase contacts with mu DNA ends. J. Biol. Chem. 266:20476–20482.

PART V

PHAGES BY HOST OR HABITAT

This page intentionally left blank

31

Viruses of Archaea

KENNETH M. STEDMAN
DAVID PRANGISHVILI
WOLFRAM ZILLIG

The Archaea (previously known as archaebacteria) were originally identified as such by Carl Woese in the late 1970s (121) and are now accepted as one of the three major divisions or domains of life, along with the Bacteria and the Eukarya. They have recently been shown by molecular techniques to be ubiquitous and may even be the dominant organisms in the open ocean (22, 45). Phylogenetically the Archaea have been split into four major phyla: the Euryarchaeota, the Crenarchaeota, the Korarchaeota, and the Nanoarchaeota. The Euryarchaeota comprise the methanogens (6), the extreme halophiles, and the extremely thermophilic *Thermococcales* (29). All isolated members of the Crenarchaeota are extreme thermophiles; however, some as yet uncultured members appear to be mesophilic or even psychrophilic, for example in marine environments (28). Korarchaeota are not yet represented by any cultured organisms (8, 56, 87). The recently discovered phylum Nanoarchaeota is represented to date by only one apparently parasitic organism (38).

In comparison with viruses of Bacteria and Eukarya, relatively little work has been done on viruses of Archaea. Most characterized viruses of Euryarchaea have typical "head and tail" morphology, similar to that of many bacteriophages, and belong to the known families *Myoviridae* and *Siphoviridae*. By contrast, all characterized viruses of Crenarchaeota have unusual morphotypes, and novel viral families had to be introduced for their classification.

As in the case of Bacteria, a number of fundamental discoveries on the nature of Archaea have been made by the study of their viruses. One was the identification of transcriptional promoter sequences in the *Sulfolobus* virus SSV1 that resembled eukaryotic promoters (82) (reviewed in 125). The discovery that the DNA packaged in SSV1 particles is positively supercoiled was one of the first examples of stably positively supercoiled DNA in nature (62). There are undoubtedly many more discoveries to follow.

Recently a number of complete sequences of archaeal viruses have been determined (5, 23, 48, 71, 73, 76, 110). Their genome sequences have few similarities to other known sequences in the sequence databases, although some of these are to sequences of bacteriophage and animal viruses (discussed below). There are striking similarities between some archaeal viral genomes, possibly in part due to horizontal gene transfer. Similar to genomes of bacteriophages, those of some viruses of the Archaea have a mosaic structure and are subject to a great deal of modification by recombination, deletion, and rearrangements (51). Comparison of complete genome sequences of archaeal hosts and viruses has allowed the identification of a number of putative cryptic proviruses (55, 95).

Some of the viruses of Archaea have served as the basis for the development of molecular genetics for the Archaea, a major bottleneck for the study of these organisms (17, 102), and some were discovered for that purpose. One, the virus ΨM1 of *Methanothermobacter marburgensis*, was shown to be a general transducing virus (58) and there is also a report of a general transducing particle for *Methanococcus voltae* (10). Another, SSV1, has been modified to be an infectious shuttle vector that replicates both in *Sulfolobus solfataricus* and in *E. coli* (102).

Four of the viruses described in this chapter, the halobacterial virus ΦH and the *Thermoproteus* viruses TTV1 to TTV3, have been extensively reviewed in the previous edition of this work (131) and little has been published on them since. Therefore, data on these viruses will only be briefly summarized here.

Viruses of Euryarchaeota

The best studied of the viruses from the Euryarchaea are the viruses ΦH from *Halobacterium salinarum* studied by Wolfram Zillig's group in the late 1980s (reviewed in 131), the virus HF2 from *Haloferax volcanii* which has recently been sequenced in its entirety by Michael Dyall-Smith's group (110), and ΦCh1 from the haloalkaliphile *Natrialba magadii* studied by Angela Witte's group and also recently sequenced (48). The *Methanothermobacter marburgensis* virus ΨM1 and its deletion variant, ΨM2, have been extensively studied by Thomas Leisinger's group and were also recently sequenced (76). Most of the viruses with euryarchaeal hosts have typical head and tail morphology and have been assigned to either the *Myoviridae* or the *Siphoviridae* virus families (http://www.ncbi.nlm.nih.gov/ICTVdb/) (for an overview of phage classification, see chapter 2).

Viruses of Halobacteria

Extreme halophiles of the class Halobacteria (31) thrive in conditions of near-saturating salt and are the dominant organisms in many hypersaline environments (69). They also include the extreme haloalkaliphiles, discussed in more detail below. All characterized halophages with the exceptions of His1 (9), and the recently described His2 and SH1 (23) (see below), have head and tail morphologies of varying sizes. All have double-stranded linear DNA genomes that vary from ca. 30 kbp (Hh3) to 230 kbp (Ja1), although the latter was not measured directly (117). Pertinent characteristics and references are listed in table 31-1 and representative morphologies are shown in figure 31-1. A recent review on novel haloarchaeal viruses is available (23).

ΦH (*Myoviridae*)

Halophage ΦH is probably the most studied virus of extreme halophiles and has been extensively reviewed (108, 131). It appeared "spontaneously", probably via prophage induction, in a laboratory strain of *H. halobium* (now *H. salinarum*) (93) in a manner very similar to that of the unrelated ΦN (116). Halophage ΦH has a typical myovirus structure with an icosahedral head and a contractile tail with tail fibers (93) (figure 31-1). The ends of the genome are terminally redundant, indicating that it replicates via a headful packaging mechanism (93). It has a highly variable genome due to recombination with its host (75) and also duplication and inversion of the so-called L region of the viral genome (90, 92). Halophage ΦH is temperate and its lysogeny is similar to that of coliphage P1 (reviewed in chapter 24) as its genome is stably maintained as a circular plasmid and does not integrate into the host genome (91).

The complete genome sequence of the host of ΦH, *H. salinarum,* confirmed the lack of an integrated prophage (64). There are three putative homologs to the phage repressor protein in the *H. salinarum* genome but no other genes indicating an interaction with the genome, for example as an integrated prophage, except the known IS elements (see http://www.halolex.mpg.de/ for an overview of *H. salinarum* and its sequence).

Repression of the prophage genome is complex. Firstly, it is regulated by a viral transcriptional repressor (109). Secondly, translation is repressed by an antisense RNA-based system whereby an antisense RNA binds to the sense T4 transcript and the 3′ single-stranded region of the coding transcript containing the ribosome binding site is removed (106). Additional RNA processing occurs in ΦH replication and is reviewed in Stolt and Zillig (108).

Table 31-1 Viruses of Halobacteria

Virus	Host	Head/tail (nm)	Genome size (kbp)	Genome G-C (%)	Lytic/ temperate	Eclipse period (h)	Latent period (h)	Salt-sensitive	Burst size	References
ΦH	*Halobacterium salinarum*	64/170c	59	64	T	5.5	7	+	170	(107, 131)
ΦN	*H. salinarum*	55/80n	56	70	N.D.	10	14	−	400	(116)
Hs1	*H. salinarum*	50/120c	N.D.	N.D.	L	12	17	+	200–300	(111, 112)
Ja1	*H. salinarum*	90/150c	230	N.D.	N.D.	2	6	+	140	(117)
Hh1	*H. salinarum*	60/100n	37.2	67	T	6	12	+	1100	(72)
Hh3	*H. salinarum*	75/50	29.4	62	T	5	8	+	425	(72)
S45	*H. salinarum*	40/70	N.D.	N.D.	L	N.D.	N.D.	+	1300	(20)
S5100	*H. salinarum*	65/76c	N.D.	N.D.	L	5.5	9	−	60–65	(19)
S50.2	*H. salinarum*	63/78c	N.D.	N.D.	L	5.5	9	−	60–65	(21)
S4100	*H. salinarum*	56/85c	N.D.	N.D.	L	5.5	9	−	60–65	(21)
S41	*H. salinarum*	89/141c	N.D.	N.D.	L	5.5	9	−	60–65	(21)
HF1	*Halobacterium* sp.	58/94	79.7	N.D.	L	N.D.	N.D.	+	N.D.	(23, 68)
HF2	*Halorubrum coriense*	58/94	77.670	55.9	L	N.D.	N.D.	+	65	(68, 110)
His1	*Haloarcula hispanica*	74 × 44/7[a]	14.462	N.D.	L	4-6	3	−	<1	(9)
His2	*Haloarcula hispanica*	62 × 69/0	16.2	N.D.	L	N.D.	1	−	9	(23)
SH1	*Haloarcula hispanica*	55[b]	27	N.D.	L	N.D.	N.D.	N.D.	N.D.	(23)
ΦCh1	*Natrialba magadii*	70/130c	54.498	61.9	T	5	11	+	150	(48, 120)

c, contractile tail; n, noncontractile tail; N.D., not determined or not reported.

[a] Dimensions of asymetrical head.

[b] Spherical virus particle.

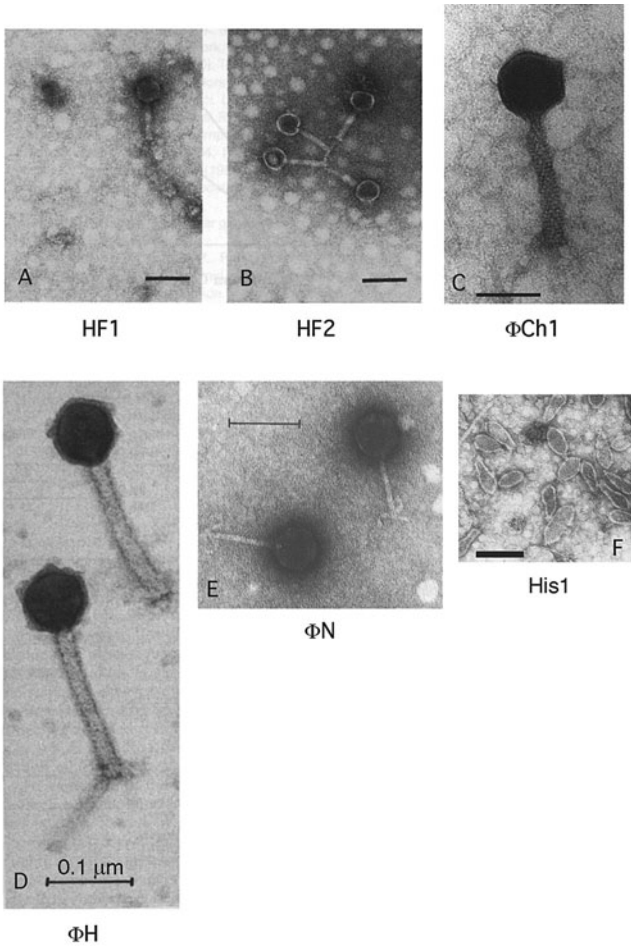


Figure 31-1 Viruses of Halobacteria. Transmission electron micrographs of viruses from extreme halophiles. All negative stain; bars represent 100 nm except 70 nm in panel C. A: HF1, B: HF2, C: ΦCh1, D: ΦH, E: ΦN, F: His1. Panels A and B reprinted from Nuttall and Dyall-Smith (68) with permission. Panel C courtesy of A. Witte. Panel D reprinted from Zillig et al. (127) with permission. Panel E reprinted from Vogelsang-Wenke and Oesterhelt (116) with permission. Panel F reprinted from Bath and Dyall-Smith (9) with permission.

HF1 and HF2 (Unclassified; *Siphoviridae*)

The haloviruses HF1 and HF2 were discovered as part of a survey of Australian salterns for lytic viruses of extreme halophiles with a wider host range than just *H. salinarum* (68). They have typical head and tail morphology (figure 31-1). As with all other known viruses of extreme halophiles, they have double-stranded linear DNA genomes (table 31-1). Unlike HF1, HF2 does not have terminal redundancy of its genome but contains 306 bp direct terminal repeats and must therefore replicate in a different manner (67).

The complete genome of HF2 was recently sequenced (110). Four tRNA genes were found in the virus genome, possibly because the viral genome has a different nucleotide distribution and codon usage compared with its host, *Halorubrum coriense*. The G-C content for *H. coriense* has not been determined; however, most Halobacteria have about 66% G-C in their genomes (31) compared with only 56% of the HF2 genome. Similar to most genomes of viruses of Archaea, there were few (ca. 10% of the open reading frames in the HF2 genome) matches to any sequences in the public databases. The significant matches to bacteriophage genes include mycobacteriophage D29, *Haemophilus* phage HP1, the terminase of the enterobacterial phage RB49. Matches to organismal genes include the Bacteria *Aquifex*, *Listeria*, *Sinorhizobium*, and *Synechocystis*, and the archaeon *Thermoplasma;* for details see Tang et al. (110). The HF2 genome also contains a DNA polymerase gene and a number of helicase genes in addition to an integrase/recombinase gene, even though the genome does not appear to integrate into the host.

The HF1 virus was isolated from the same source as HF2. The two viruses are morphologically identical (figure 31-1) and have very similar genome size and protein composition. However, the two viruses have different host ranges (table 31-1), plaque morphologies, and sensitivity to chloroform (68). The recently reported genome sequence of HF1 (23) was found to be a recombinant between HF2 and a novel, related, but as yet undiscovered virus. About 60% of the genomes of HF1 and HF2 are identical whereas the rest of the sequence is only 87% identical (23).

ΦCh1 (*Myoviridae*)

The virus ΦCh1 was found by spontaneous lysis of a culture of the haloalkaliphilic archaeon *Natrialba magadii*, an isolate from the soda lake Lake Magadii in Kenya. This organism grows optimally at 3.5 M NaCl and pH 9.5. Unlike all other known viruses of Archaea, it contains both host RNA and viral DNA in its virion (120). The DNA genome, like those of all other viruses of extreme halophiles, is double-stranded and linear. The genome of ΦCh1 was found to integrate into the host genome (120). Some of the DNA in the virus genome was found to be methylated by a virus-encoded methylase (7). Despite the low salt concentration of the *E. coli* cytoplasm, this methylase was able to complement *dam* mutants of *E. coli* (7). Surprisingly, the sequence of the main coat protein of ΦCh1 was found to be similar to the sequence of a coat protein from ΦH, even though the hosts of these two organisms thrive in very different environments (49).

The recently determined complete genome sequence of ΦCh1 was found to contain further sequences similar to the partially sequenced ΦH genome (33, 48). In some regions the nucleotide sequence identity was as high as 97%, indicating potential horizontal gene transfer (48). However, ΦCh1 is not able to infect *H. salinarum* (120). Eighty of the 98 open reading frames (ORFs) in the ΦCh1 genome did not show any similarity to genes of known function, similar to HF2 (see above) and the other viruses of Archaea whose genomes have been sequenced (see below). There are two putative integrase genes in the ΦCh1 genome, although it is unclear which, if any, of them is used for the integration of the provirus. Not only was the previously known methyltransferase gene of ΦCh1 (7) identified but two more putative methyltransferase genes were also found, including one 57% identical to the ΦH cytosine methyltransferase gene (105). One ORF showed highly significant similarity to the PCNA protein, the processivity factor for archaeal and eukaryotic replicative DNA polymerases (41). It will be interesting to see whether it modifies the activity of the *N. magadii* DNA polymerase or a viral polymerase that was not identified. The genome of ΦCh1 otherwise is organized in a manner resembling that of "early," "middle" and "late" genes of tailed bacteriophages (16), indicating functional modularity.

Other Viruses of Halobacteria

Six different viruses of *H. salinarum* (previously known as *H. cutirubrum*) have been isolated from Jamaican salt ponds: Ja1, S45, S5100, S50.2, S4100, and S41. They seem to differ mostly in their mechanism of host attachment (21). Their infectivity depends on the salt concentration of their environment, like that of one of the first discovered viruses of extreme halophiles, Hs1 (111, 112). The response to salt concentration is probably important in the environment as salt ponds are often diluted rapidly by rainfall such that the obligately halophilic hosts do not survive. The *H. salinarum* virus ΦN also is able to maintain 50% of its infectivity after 14 hours in distilled water (116). Its genome sequence differs from that of ΦH, as indicated by Southern hybridization (116). Its genome is fully cytosine methylated (116).

The only characterized viruses of extreme halophiles that do not have a "head and tail" bacteriophage-like morphology are His1 and His2 from salterns and salt lakes in Australia (9) and the recently reported SH1 (23). Like ΦN, His1 is resistant to low salt concentrations (9). His1 has similar morphology and genome size to that of the well-studied fusellovirus SSV1 of the thermoacidophilic crenarchaeon *Sulfolobus*. However, it is lytic, not temperate, does not integrate into the host genome, has a linear rather than circular genome, its replication is terminally protein primed, and there is no sequence similarity to that of SSV1 or other members of the *Fuselloviridae* (23, 71). Therefore, His1 should not be considered a member of the *Fuselloviridae* (23) and a new proposal for classification of this virus is pending with the International Committee on the Taxonomy of Viruses (ICTV) (M. Dyall-Smith, personal communication). SH1 similarly is a lytic double-stranded DNA virus with a linear genome; particles are spherical and possess a lipid layer underneath an outer protein layer (23).

Virus-like Particles in Hypersaline Environments

Direct electron microscopic observation of samples from the Dead Sea after a bloom of halophilic Archaea revealed many virus-like particles with similar morphology to His1 and His2, as well as particles with an unusual star-shaped morphology (70). Attempts to cultivate the halobacterial host strains of these particles were unsuccessful (70). The dominant halobacterial species in these environments as determined by molecular techniques has also not been cultivated (35). Similar "orphan" spindle-shaped virus-like particles have been obtained from samples from solar salterns (34).

Viruses of Methanogens

There are considerably fewer characterized viruses of methanogens than of either Halobacteria or Crenarchaea (table 31-2). It is unclear whether this is due to the absence

Table 31-2 Viruses of Methanogens

Virus	Host	Head/tail (nm)	Genome size (kbp)	Genome F-C content (%)	Lytic/ temperate	Eclipse period (h)	Latent period (h)	Burst size	References
ΨM1	*Methanothermobacter marburgensis*	55/210	30.4	N.D.	L	N.D.	4	8	(59)
ΨM2	*M. marburgensis*	55/210	26.11	46.3	L	N.D.	N.D.	N.D.	(44)
ΨM100	*M. wolfeii*	pp	28.79	45.4	pp	N.D.	N.D.	N.D.	(53)
ΦF1	*M. thermoautotrophicus*	70/160 × 20n	85	N.D.	L	N.D.	N.D.	N.D.	(66)
ΦF3	*M. thermoautotrophicus*	55/230 × 9n	36	N.D.	L	N.D.	N.D.	N.D.	(66)
PG	*Methanobrevibacter smithii*	N.D.	ca. 50	N.D.	L	N.D.	7–9	20	(11)
PMS11	*Methanobrevibacter smithii*	N.D.	35	N.D.	L	N.D.	N.D.	N.D.	(50)
VLP	*Methanococcus voltae* A3	52 × 70	23	N.D.	T	N.D.	N.D.	N.D.	(123)
VTA	*M. voltae* PS	40/61	4.4	N.D.	N.D.	N.D.	N.D.	N.D.	(24)

pp, prophage; n, noncontractile; N.D., not determined or not reported.

of viruses or insufficient screening. The relative difficulty of plating the methanogens as lawns (114) may have contributed to this situation. There are, however, many plasmids of methanogens which have been used to develop molecular genetic tools (115), potentially lessening the interest of screening for viruses. Two viruses of methanogens have only been mentioned in meeting abstracts and it is unclear why they were not further investigated (11, 50).

ΨM1, ΨM2 (*Siphoviridae*)

The best studied viruses of methanogens are ΨM1 and its deletion variant ΨM2 that infect *Methanothermobacter marburgensis* (formerly known as *Methanobacterium thermoautotrophicum;* 118). The complete genome of ΨM2 has been sequenced (76). ΨM1 is a lytic virus that was isolated from an anaerobic sludge digester at a temperature of 55–60 °C (59). The virus particle (figure 31-2) has an icosahedral head and a flexible tail (59). In addition to its approximately 30.4 kbp linear double-stranded DNA genome, virions contained multimers of a cryptic plasmid from *Methanothermobacter*, pME2000 (59), that has no similarity in sequence to ΨM1. The packaged genome of ΨM1 was found to be circularly permuted and to have an approximately 3 kbp terminal redundancy at both ends (44). Both these facts led to the proposition that the genome is packaged by a headful mechanism. ΨM1 was shown also to transduce a number of genetic markers (58), and is the only known transducing virus for any of the Archaea.

The only known host of ΨM1 is *Methanothermobacter marburgensis* and the viral genome did not hybridize to the genomic DNA of the uninfected host. There was significant hybridization, however, to the genomic DNA of *Methanobacterium wolfeii* (59), due to sequence similarity to its defective prophage ΨM100. Resistance of *M. wolfeii* to infection with ΨM1 is probably due to superinfection immunity also provided by the defective prophage (53). The lysate produced by ΨM1 infected *M. marburgensis* cells was, however, still able to lyse both *M. wolfeii* and *M. marburgensis* by the activity of a psuedomurein endoisopeptidase that is encoded by the viral genome (76, 101).

The totally sequenced genome of ΨM2 was found to contain 26,111 bp with a G-C content of 46.3%, slightly lower than that of *M. marburgensis* (76, 113). The genome does not encode any tRNA genes and the codon usage for the virus is similar to that of its host (76). Of 31 ORFs greater than 90 amino acids in length, only six were functionally assigned. This percentage of identified ORFs is similar to those of the viruses HF2 and ΦCh1 of the extreme halophiles. Many ORFs appear to be cotranscribed and genes with putative functions are clustered, again as in bacteriophage and halovirus genomes (33, 48, 76). There were ORFs similar to bacteriophage genes. Four of the ΨM2 ORFs were similar to genes from *Bacillus subtilis* phage PBSX, including a putative terminase gene and two structural protein genes, one of which was shown to be a structural protein of ΨM2 (76). There was also a putative portal gene with sequence similarity to the portal gene of phage HP1 of

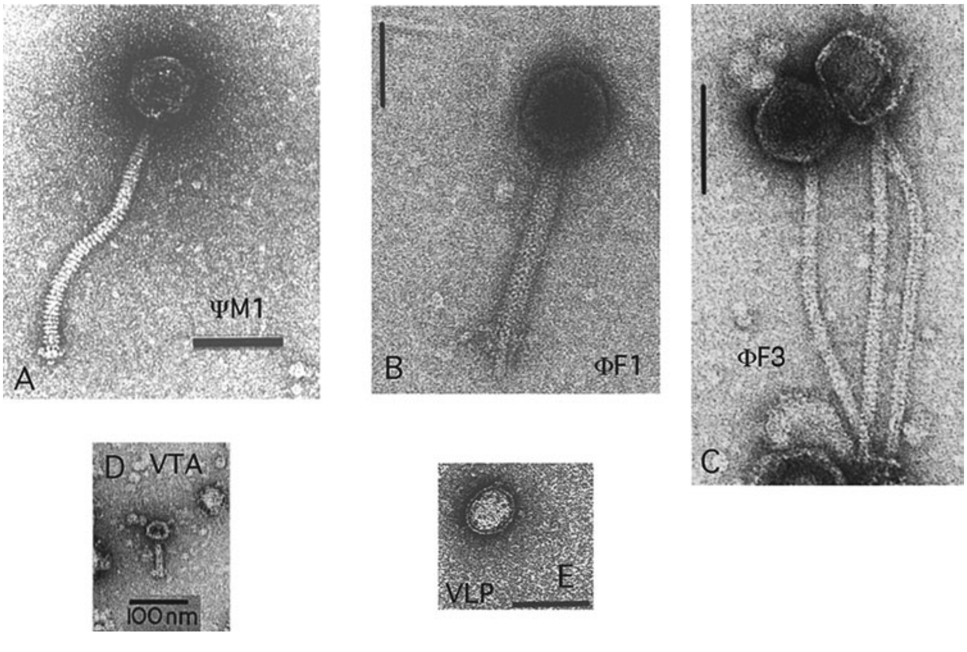


Figure 31-2 Viruses of methanogens. Transmission electron micrographs of viruses from methanogens. All negative stain; bars represent 70 nm, except VLP and VTA 100 nm. A: ΨM1, B: ΦF1, C: ΦF3, D: VTA, E: VLP. Panel A from Meile et al. (59) with permission. Panels B and C from Nolling et al. (66) with permission. Panel D from Eiserling et al. (24) with permission. Panel E from Wood et al. (123) with permission.

Haemophilus influenzae (25). Finally, the putative tail protein gene showed similarity to the tail protein gene from phage L5 of *Mycobacterium tuberculosis* (36, 76). Surprisingly, as in HF2 an integrase/recombinase gene of the phage λ family was also found (for review of phage λ biology, see chapters 8, 9 and 27). It is intriguing that both HF2 and ΨM2 that appear to be obligately lytic yet both contain putative integrase genes, like the known integrative virus ΦCh1. However, these gene products may be acting as recombinases instead of integrases, like phage XerC in *H. influenzae* (26, 76).

ΨM100 (Unclassified; *Siphoviridae*)

The deletion that took place in the generation of ΨM2 from ΨM1 was found to remove a 692 bp fragment between two 82 bp direct repeats (76). This is similar to a deletion of an 8 kbp fragment between direct repeats of 85 bp in the conjugative plasmid pNOB8 of the thermophilic crenarchaeon *Sulfolobus* (98). The mechanisms for these deletions are unknown.

The defective prophage present in the *M. wolfeii* genome, ΨM100 (104), was found to be 28,798 bp in length, slightly larger than the ΨM2 genome (53). Almost all of the difference is made up of an insertion of 2793 bp with anomalously low G-C content (53). The ORFs in this insertion were similar to ORFs from the *Methanothermobacter autotrophicus* ΔH genome (100). Otherwise the ΨM100 prophage genome has 71% overall nucleotide sequence identity to ΨM2 (53). The putative packaging (*pac*) site and origin of replication for ΨM2 were more than 98% identical to those of ΨM100. However, this is also the region where the insertion in the ΨM100 genome is present, so the replication defect in ΨM100 may be due to the insertion. All but four ORFs in the ΨM100 genome are from 40% to 100% identical in predicted amino acid sequence to ΨM2 ORFs. Surprisingly, the pseudomurein endoisopeptidase encoded by ΨM100 is one of the least similar (54). There was one ORF that was missing and one extra one in the ΨM100 genome. One ORF was very similar in its N-terminus in the two genomes but appeared to have insertions in the C-terminus in ΨM100. Since the genomes of ΨM100 and ΨM2 are so similar it is surprising that no virus particles have ever been observed in *M. wolfeii* cultures (59, 104). Flanking the prophage genome in the host are 21 bp direct repeats of all adenine and thymine residues that are probably the attachment sites (53). Whether the virally encoded integrase can act on these sites awaits experimental confirmation.

Other Viruses of Methanogens

Other viruses of methanogens have been considerably less well characterized (see table 31-2). A DNase A-resistant DNA was found in filtered cultures of *Methanococcus voltae* strain PS that was also able to transduce a number of markers (10). A small virus-like particle was observed in these cultures but it was of very low titer and could not be induced by ultraviolet irradiation or treatment with mitomycin C (24). This *voltae* transfer agent (VTA) appeared to carry only about 4400 bp of apparently random DNA (10). The transducing activity and the virus-like particles were also degraded rapidly both aerobically and anaerobically (10).

Two lytic viruses of *Methanothermobacter thermoautotrophicus* (previously known as *Methanobacterium thermoformicicum*), ΦF1 and ΦF3, were discovered in a 55 °C experimental sludge-bed reactor (66). Both had typical head and tail morphologies with apparently noncontractile tails (see figure 31-2). They both had double-stranded DNA genomes; ΦF1 had a linear genome of about 85 kbp in size, ΦF3 of about 36 kbp in size. The ΦF3 genome is either linear with terminal redundancy or circular according to the physical map of the genome (66). ΦF1 had a much wider host range than ΦF3, including *Methanothermobacter thermoautotrophicus* ΔH, one of the strains for which a complete genome sequence is available (100). The host range of ΦF1 appeared to be related to the restriction/modification system encoded by the host strains as there was a strong selection against the CTAG sequence in the ΦF1 genome (65). The genomes of ΦF1 and ΦF3 did not hybridize to each other, to the host chromosomal DNA or to DNA from the previously characterized methanogen virus, ΨM1 (see above) (66).

A virus-like particle was found in cultures of *Methanococcus voltae* strain A3 and isolated from unwashed cells (123). The morphology was similar to that of SSV1 of *Sulfolobus* (see below), although more heterogeneous (see figure 31-2). The double-stranded circular DNA of this particle was 23 kbp, considerably larger in size than that of SSV1. This DNA was also found to be integrated into the host genome (123). The particles appeared to have an envelope of one major protein species, unlike SSV1 (123). No induction or infectivity of the particles could be shown. No sequence data are available for the DNA of these virus-like particles.

Viruses of Thermococcales

The *Thermococcales*, including *Pyrococcus*, comprise some of the most thermophilic organisms. Many different isolates have been obtained, mostly from marine hydrothermal vents, both shallow and from the deep sea (132). There has been considerable study of enzymes from *Pyrococcus* species for biotechnology applications (40). The genomes of three *Pyrococcus* species—*P. horikoshii*, *P. abyssi*, and *P. furiosus*—have been sequenced (see 135). However, the development of genetic tools for these organisms has been relatively slow (52). Virus-like particles have been observed in *Pyrococcus* enrichment cultures from samples from deep-sea hydrothermal vents. (W. Zillig and I. Holz, personal communication; P. Forterre and E. Marguet, personal communication; 30a). One head-and-tail virus from *P. abyssi* has been isolated and its genome has been sequenced (30b). There have also been reports of cryptic prophage in *Pyrococcus* genomes. One 30 nm icosahedreal virus-like particle of *Pyrococcus woeseii* was detected after spontaneous lysis of the strain on entry to the stationary growth phase (128) but has not been further studied.

Viruses of the Crenarchaeota

The first viruses from Crenarchaeota have been characterized by Wolfram Zillig and coworkers. The cultivated viruses infect extremely thermophilic members of the genera *Thermoproteus*, *Sulfolobus*, and *Acidianus*. Many virus-like particles have been observed in enrichment cultures from samples from environments dominated by Crenarchaea. Unlike viruses of the Euryarchaeota, none of these viruses or virus-like particles have been shown to have head-and-tail bacteriophage-like morphology. The unique morphology, unusual genome structures, and genome sequences required the introduction of three novel virus families, *Fuselloviridae*, *Lipothrixviridae*, and *Rudiviridae*, by the ICTV. One additional family, *Guttaviridae*, has been proposed (4). Other novel viruses (see below) may necessitate the creation of yet another family (table 31-3; figure 31-3).

Viruses of Thermoproteus tenax

Thermoproteus tenax is a facultatively chemolithotropic strictly anaerobic crenarchaeote originally found in a mud hole in Iceland that grows optimally at 88°C (133). It can grow autotrophically by the reduction of elemental sulfur to hydrogen sulfide or by sulfur respiration (133). The metabolism of *T. tenax* has been well studied (99) and its genome is being sequenced (B. Siebers and R. Hensel, personal communication). Three different viruses of *T. tenax*—TTV1, TTV2, and TTV3—were found in the Kra1 strain from the Krafla volcano, Iceland on transfer from heterotropic to autotrophic growth conditions (42). A fourth, TTV4, was found in a fresh sample from the Krafla volcano and lysed all enrichment cultures (131). TTV1, TTV2, and TTV3 have linear double-stranded DNA genomes, are unrelated according to hybridization analysis, and have different protein composition.

TTV1 (*Lipothrixviridae*)

The best studied of these viruses is TTV1 and it has been extensively reviewed (131). The virions of TTV1 are flexible rods (figure 31-3) with a central core enclosed in an envelope that contains host lipids (131). The virion consists of at least four proteins: two DNA binding proteins, one envelope protein, and a fourth protein whose location is unknown. About 80% of the genome has been sequenced (EMBL accession X14855). The nature of its ends is presently unclear.

TTV1 persists in a "carrier state" in its host (42). Occasionally some recombination with the host genome was observed (131). The genome of TTV1 varies a great deal due to small insertions and deletions (70). The insertions do not appear to be IS elements as they are too small and do not have the inverted or direct repeats typical of these structures (63). The insertions have very different G-C content from the rest of the TTV1 genome, and it is unknown whether they are present in the host chromosome.

Table 31-3 Viruses of the Crenarchaeota

Virus	Host	Virion dimensions (nm)	Genome size (kbp)	genome G-C content (%)	Lytic/ temperate	Family	References
TTV1	*Thermoproteus tenax*	40/400	15.9	37.0	T	*Lipothrixviridae*	(42)
TTV2	*T. tenax*	20/1250	16	N.D.	T	*Lipothrixviridae*	(42)
TTV3	*T. tenax*	30/2500	27	N.D.	N.D.	*Lipothrixviridae*	(42)
TTV4	*T. tenax*	30/500	17	N.D.	L	Unassigned	(131)
DAFV	*Acidianus ambivalens*	2200/27	56	N.D.	T	*Lipothrixviridae*	(129)
AFV	*Acidianus* sp.	900/24	20.1	N.D.	N.D.	*"Lipothrixviridae"*	(12)
SSV1	*Sulfolobus shibatae*	60 × 90	15.495	39.7	T	*Fuselloviridae*	(57)
SSV2	*"S. islandicus"*	55 × 80	14.794	38.5	T	*Fuselloviridae*	(2, 103)
SSV3	*"S. islandicus"*	55 × 80	15	N.D.	T	*Fuselloviridae*	(126)
SSVK1	*Sulfolobus* sp.	N.D.	17.779	38.8	T	*Fuselloviridae*	(88)
SSVY1	*Sulfolobus* sp.	N.D.	16.473	38.8	T	*Fuselloviridae*	(88)
SIRV1	*"S. islandicus"*	780/23	32.312	25.2	T	*Rudiviridae*	(77)
SIRV2	*"S. islandicus"*	900/23	35.502	25.3	T	*Rudiviridae*	(77)
SIFV	*"S. islandicus"*	1950/24	41.050	33.3	T	*Lipothrixviridae*	(5)
SNDV	*S. neozealandicus*	110–185/95–75	30	N.D.	T	"*Guttaviridae*"	(4)
STIV	*Sulfolobus* sp.	70	19	N.D.	T	Unassigned	(88)

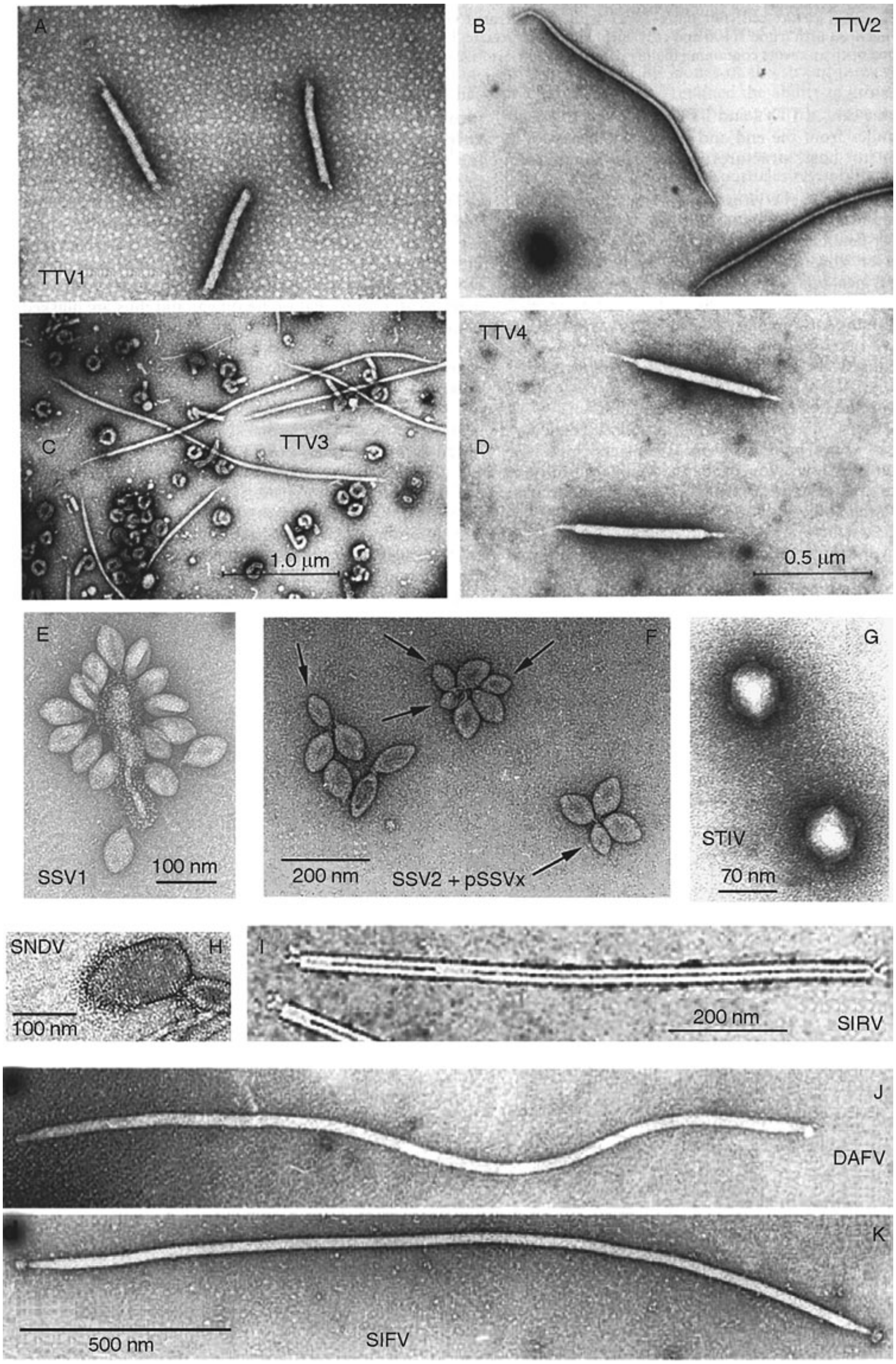

A
TTV1
B
TTV2
C
TTV3
1.0 μm
TTV4
D
0.5 μm
E
SSV1
100 nm
F
200 nm
SSV2 + pSSVx
G
STIV
70 nm
SNDV
H
100 nm
I
200 nm
SIRV
J
DAFV
K
500 nm
SIFV

Figure 31-3 Viruses of the *Crenarchaeota*. Transmission electron micrographs of viruses from crenarchaeal hosts. All negative stain; scale bars as labeled in panels A: TTV1, B: TTV2, C: TTV3, D: TTV4, E: SSV1, F: SSV2 plus pSSVx, G: STIV, H: SNDV, I: SIRV2, J: DAFV, K: SIFV. Panels A to D reprinted from Zillig et al. (127) with permission. Panel E reprinted from Stedman et al. (102) with permission. Panel F reprinted from Arnold et al. (2) with permission. Panel G, K.M.S. unpublished. Panels H–K reprinted from Zillig et al. (126) with permission.

TTV2 and TTV3 (*Lipothrixviridae*)

TTV2 and TTV3 were not observed as frequently as TTV1 in cultures of the Kra1 strain and differ in their flexibility and length from TTV1 and TTV4. All tested subclones of *T. tenax* Kra1 produced TTV2 but production of the virus was usually low and only occasionally strongly induced (131). The large TTV3 has been observed even less frequently than TTV2 and very little is know about it (131). The two viruses are similar in morphology to the *Sulfolobus* virus SIFV and to the *Acidianus* virus DAFV (see below).

TTV4 (Unclassified)

The virions of TTV4 are stiff rods (figure 31-3). TTV4 appears to be one of the most resistant viruses known, as it is stable and infectious after 1 hour of autoclaving at 120 °C (3). TTV4 is also particularly virulent and spreads easily, complicating its study (131).

Viruses of Acidianus

Members of the archaeal genus *Acidianus (Desulfurolobus)* are thermophilic acidophiles that can grow both heterotrophically (*A. brierleyi*) or autotrophically (*A. brierleyi, A. ambivalens,* and *A. infernus*) either aerobically by oxidizing sulfur to sulfuric acid or anaerobically by reducing sulfur to hydrogen sulfide (94, 134). They are commonly found in autotrophic enrichment cultures from acidic hot springs throughout the world. The genus groups with the *Sulfolobus* species in the *Sulfolobaceae* family (39). In addition to two viruses of *Acidianus* listed below, a number of virus-like particles produced by different *Acidianus* strains have recently been found in samples and enrichment cultures from Yellowstone National Park in Wyoming, USA (YNP) (81).

DAFV (*Lipothrixviridae*)

A large filamentous virus about 2.2 μm long and 27 nm wide was found in an *Acidianus* isolate from Iceland (figure 31-3) (129). This virus, DAFV (*Desulfurolobus ambivalens* filamentous virus), had very similar morphology to TTV2 and SIFV (figure 31-3) and virus production could be induced by ultraviolet irradiation. Surprisingly it appeared to be able to infect a *Sulfolobus* strain in a coinfection with SIRV1 (129). Otherwise it resembled SIFV in both protein profile and lipid content and showed some DNA cross-hybridization (5). No sequence data are available for DAFV.

AFV (‘*Lipothrixviridae*’)

AFV1 (*Acidianus* filamentous virus 1) was produced by an *Acidianus* strain isolated from a hot spring in YNP (12). Filamentous virions, 900 nm long and 24 nm wide, are covered with a lipid envelope and contain at least five different proteins with molecular masses from 23 to 130 kDa. With the help of unusual claw-like termini the virions attach to pili of host cells. The host range is confined to several *Acidianus* isolates from YNP. The genome is a 20.1 kbp long, linear double-stranded DNA that has been sequenced. Of 40 ORFs longer than 48 amino acids, 12 are similar to ORFs present in Rudiviruses and Lipothrixviruses from *Sulfolobus*. One ORF is similar to an ORF present in Fuselloviruses (12).

Viruses of Sulfolobus

Many viruses of *Sulfolobus* have been studied, mainly because of their relative ease of isolation (129). *Sulfolobus* grows optimally at 80 °C and pH 3 and is an obligate aerobe (32). It grows well on solid media both as single colonies and in lawns (129). The complete genome of *Sulfolobus solfataricus* strain P2 has recently been sequenced (97). Viruses of *Sulfolobus* have also recently been reviewed (79, 126, 130). Several viruses of *Sulfolobus* have had their complete or nearly complete genome sequenced (see below).

SSV1 (*Fuselloviridae*)

The first discovered and best studied of the viruses of *Sulfolobus* is SSV1 (*Sulfolobus shibatae* virus 1) (reviewed in 3, 126). SSV1 DNA was originally found as a plasmid in *S. shibatae* from Beppu Onsen, Japan (124). Later it was shown to be the genome of a UV- inducible virus-like particle (57). Finally, by showing infectivity for *S. solfataricus*, SSV1 was shown to be a virus (89). The virus particle has an unusual spindle shape (figure 31-4) with a short tail at one end. The SSV1 genome is present in infected cells both as an episomal plasmid and as an integrated provirus (124). In virus particles it is packaged in a positively supercoiled state (62). The viral genome integrates specifically into an arginyl-tRNA gene of the host (85). This integration disrupts the viral integrase gene via recombination in a 44 bp segment that duplicates the 3′ end of the tRNA gene. Therefore integration does not disrupt the tRNA gene. In vitro the viral integrase gene product is necessary and sufficient for both the integrative and excisive reactions with oligonucleotide templates (60, 61).

SSV1 has two highly hydrophobic coat proteins, VP1 and VP3, the genes of which contain a directly repeated DNA sequence (84). The virions appear to be packaged in infected host cells in regions where the cellular membrane is replaced by islands of viral coat proteins that, in contrast to the normal cell membrane, have no contact with the S-layer. Virus particles are formed by budding through these islands (57). The virus particles themselves are very stable to high temperature (97 °C) and low pH^+ (2.0), but the packaged DNA seems to be less so (3). Virus particles contain a very basic protein, VP2, that binds DNA. The genome of SSV1 was the first of an archaeal virus to be completely sequenced

(71). Apart from the viral integrase gene, which belongs to a large family of tyrosine recombinases including proteins from halophages HF2, ΦCh1, and ΨM2, none of the ORFs matched any known proteins (71). ORFs containing cysteine codons were present in only about one contiguous half of the viral genome that contains two transcription units, leading to the proposition that the virus is the product of a fusion of two modules (71).

Analysis of the transcriptional promoter sequences led to the discovery that promoters of Archaea are similar to eukaryotic promoters (86). Replication is initiated by induction of a short transcript, T_{ind}, followed by induction of T5 and T6 transcripts (1 hour post-infection) and by proliferation of the virus (about 4 hours post-infection). However, the other transcripts in the virus genome are constitutively expressed (86). Some of these viral promoters have been used in in vitro tests for promoter function and the T6 promoter was shown to be the strongest in vitro promoter known (80).

Progress has been made in disrupting a number of genes in the SSV1 genome by a serial-selection technique (102). This technique also allowed the construction of a viral shuttle vector for *Sulfolobus* and *E. coli* that is very promising for future research. It has been used to complement both auxotrophic mutants and a metabolic mutant (43). It has even been used to functionally express a protein from the euryarchaeon *Pyrococcus furiosus* (K.M.S. unpublished data). The putative origin of the replication of SSV1 has also been used as the basis for other shuttle vectors for *S. solfataricus* and *E. coli* (17).

SSV2 (*Fuselloviridae*)

A virus named SSV2 with similar morphology and genome size to SSV1 was found in a *Sulfolobus* isolate from Reykjanes in Iceland (103). Its host was also found to produce a smaller virus-like particle that contained a cryptic plasmid, pSSVx of the pRN family of plasmids of *Sulfolobus* (2, 46, 47). Two ORFs of this plasmid are homologous to ORFs present in SSV1 and SSV2 and are presumably involved in packaging of this plasmid into the small virus-like particles (2). The SSV2 genome is 55% identical to the SSV1 genome at the nucleotide level. Nevertheless, most of the ORFs were arranged in the same order in either genome, clearly indicating that the viruses belong to the same family. The ORF identity ranged from unrecognizable to 76% (103). The integrase gene of SSV2 is conserved, as are genes encoding the viral coat proteins VP1 and VP3. However, the VP2 protein is not present in the SSV2 genome. The integrase gene of SSV2 carries a part of a glycyl-tRNA gene rather than of the arginyl-tRNA gene (103), and accordingly the SSV2 genome is integrated into a glycyl-tRNA when *S. solfataricus* is infected with SSV2 (Q. She, personal communication).

ORFs that were previously shown to be important for virus functions of SSV1 (102) were well conserved, whereas those that were less important were less conserved. Otherwise the genome is mosaic relative to SSV1, with a number of ORFs present in one genome and lacking in the other (103). By contrast to SSV1, SSV2 is induced when its host reaches the stationary phase (1). Like SSV1, SSV2 has the cysteine codons clustered in one half of the genome, adding further support to the hypothesis that SSVs were formed by module fusion similar to some bacteriophages (37, 71).

Other Fuselloviridae

Viruses similar in morphology and genome size to SSV1 and SSV2 were found in about 8% of samples taken from Icelandic solfataric fields (126). The genomes of several of these have been sequenced: SSV-Ic4 and SSV-Ic5 (X. Peng and R. A. Garrett, unpublished data). Four other SSVs have been found in YNP (SSV-Y1, SSV-Y2, and SSV-Y3) and Kamchatka, Russia (SSV-K1) (88).

The SSV-K1 genome, with 17,384 bp, is the largest SSV genome known so far (119). The genome organization is similar to those of SSV1 and SSV2, with the exception of a large insert downstream of the viral integrase gene. The viral genome integrates in another host tRNA gene, an aspartyl-tRNA gene. Like SSV2 it is also missing a VP2 gene, and like SSV1 it has been made into a shuttle vector for *Sulfolobus* and *E. coli* (K.M.S. unpublished data). The shape of this virus is more elongated than that of other SSVs.

A homolog of the VP2 gene is also missing from the genome of SSV-Y1, which was found to integrate into yet another tRNA gene of the host, a lysyl-tRNA gene (119). The SSV-Y1 genome is also mosaic in that some ORFs are well conserved with other SSVs and others are not.

In comparing sequences from all four genomes, there is no correlation between genetic and geographic divergence (119). Similar lack of correlation was obtained with partial sequences from three other SSVs from YNP (88). Whether different SSVs can simultaneously infect *Sulfolobus* or whether there is immunity to superinfection is unknown, although the integration sites do not overlap.

SIFV (*Lipothrixviridae*)

Sulfolobus islandicus filamentous virus, SIFV, was originally isolated from an Icelandic solfatara in a screen for extrachromosomal elements of *Sulfolobus* (129). SIFV has a filamentous structure. The filament is almost twice as long as the diameter of its host cell, *Sulfolobus islandicus* strain HVE10/4 (figure 31-3). It has an envelope that contains host lipids but in a different proportion to the host (5), possibly reflecting specific envelope formation. Inside the 4 nm lipid envelope is a core made up of the DNA genome and the two major virion proteins (5). From electron tomography and image reconstruction, a model of the core was derived with the proteins on the inside and the DNA wrapped around the outside, reminiscent of chromatin (5). SIFV contains spider-like tail fibers in a mop-like arrangement at its termini (5). Its host range is limited to a few closely related *S. islandicus* strains (5).

The SIFV genome does not integrate into the host genome but is present in a carrier state and is easily lost after culture dilution (5). It has been sequenced with the exception of the ends, the nature of which is unclear (5). The 40,266 bp sequenced contained 74 nonoverlapping ORFs. Similar to all other viruses of Archaea, only a few of these matched known sequences. There were also no matches with the partially sequenced genome of another *Lipothrixvirus*, TTV1, making the assignment to the same virus family questionable. However, seven ORFs had homologs in the genome of the lipothrixvirus AFV1 of *Acidianus*, and 14 in the genomes of the rudiviruses SIRV1 and SIRV2 of *Sulfolobus*. One ORF, a291, had limited similarity to an ORF from SSV1 shown by disruption to be critical for virus function (102) and is conserved in SSV2 (103) and other SSVs (119). There are three ORFs that might encode glycosyltransferases involved in the production of the virus envelope (5). There were no matches to bacteriophage genomes. Virus-like particles similar to SIFV but of different lengths have been observed in enrichment cultures from hydrothermal samples from YNP (88).

SIRV1 and SIRV2 (*Rudiviridae*)

The two known species of rudiviruses, SIRV1 and SIRV2 (*Sulfolobus islandicus* rod-shaped virus), are similar in structure and have very similar genomes. Both are maintained without change in their original hosts, but differ strikingly in infection of other hosts (77). While SIRV2 multiplies without change in several hosts, the genome of SIRV1 is subject to striking sequence variation. Genome variation is mainly caused by accumulation of point mutations, with a rate of about 10^{-3} substitutions per nucleotide per replication cycle—unprecedented for DNA viruses (77). This eventually leads to the selection of conditionally stable virus variants, coinciding with the recovery of high-fidelity replication. Such stable variants of SIRV1 produce further variants when infecting a new host, demonstrating that viral genome stability in a host does not exclude the potential to vary in other hosts (77). Mechanisms and controls underlying regulation of replication fidelity are as yet unclear.

The genomes of SIRV1 and SIRV2 have been completely sequenced (73). They are linear, covalently closed at their ends, 32.3 and 35.5 kbp long, and contain 2.0 and 1.6 kbp long inverted terminal repeats, respectively. The genomes contain 45 and 54 ORFs, respectively, of which 44 are homologous to each other. Their functions include: a dUTPase (functionally expressed in *E. coli;* 78) and a Holliday junction resolvase (functionally expressed in *E. coli;* 13). Predicted functions include two glycosysltransferases (73), a methyltransferase, a helicase, and a RecB family exonuclease (E. Koonin, personal communication).

Comparison of nucleotide sequences of the two species of rudiviruses indicates that recombination, gene duplication, horizontal gene transfer, and substitution of viral genes by homologous host genes have contributed to their evolution (73). About 15% of the ORFs of both rudiviruses have homologs in the host chromosome, about 30% have homologs in the genome of the lipothrixvirus SIFV, and about 20% in the genome of the lipothrixvirus AFV1, suggesting that the two virus families form a superfamily (73). One ORF each of the rudiviruses has a homolog in the fusellovirus SSV1. No significant sequence similarities, however, were found in genomes of viruses from euryarchaeotes.

The rudiviruses share characteristics of the linear organization of their double-stranded DNA genomes—namely covalently closed ends, long inverted terminal repeats (ITRs), and the presence in terminal regions of direct repeats and hot spots for recombination—with large eukaryal double-stranded DNA viruses including poxviruses, African swine fever virus, and *Chlorella* viruses (14, 73). Similarities between these viruses also include modes of replication and resolution of replicative intermediates. Head-to-head and tail-to-tail linked replicative intermediates of the SIRV1 genome were found, indicating that replication follows the model suggested for replication of similarly organized genomes of eukaryal viruses. According to this model, replication is initiated by the introduction of a terminal nick followed by unpairing of one strand and elongation of the 3′-OH termini of the other. On reaching the end of the template, the elongated strand folds back on itself and copies the remainder of the genome. In both virus families concatemeric replicative intermediates apparently are resolved by virus-encoded Holliday junction resolvases (14). Moreover, recognition sequences for viral resolvases in the genomes of poxviruses and rudiviruses appear to be similar (13). All these similarities might indicate evolutionary relationships between genomes of rudiviruses and some large eukaryal double-stranded DNA viruses.

SNDV (“*Guttaviridae*”)

A novel virus, SNDV (*Sulfolobus neozealandicus* droplet-shaped virus), was found in a sample from the Steaming Hill solfataric field in New Zealand. Like other viruses from Crenarchaea it has an unusual morphology, in this case of a “bearded” droplet 110–185 nm long by 75–90 nm wide with a bundle of many thin fibers on the pointed half of the virion (4; table 31-3, figure 31-3). Despite its droplet form, the viroid appeared to have helical symmetry, somewhat like the nucleocapsid of the human lentivirus HIV-1 (30). A novel virus family, “*Guttaviridae*” has been proposed for classification of the virus. The host range of SNDV is confined to several *Sulfolobus* isolates from New Zealand. Host cells could be cured by repeated dilution, indicating that the virus is present in a carrier state (4). It contains a 20 kbp

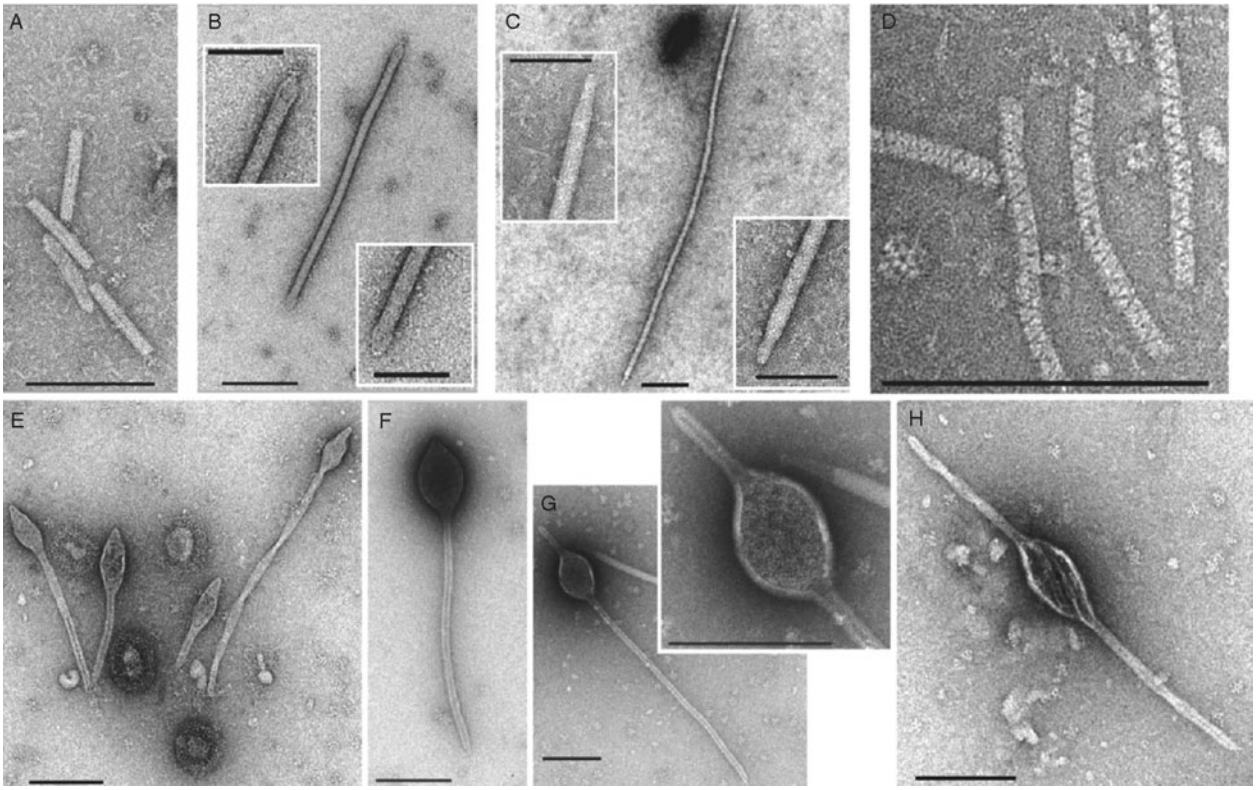


Figure 31-4 Electron micrographs of some viruses and virus-like particles from hot springs in Yellowstone National Park, USA. Bars represent 200 nm (100 nm in insets). Reproduced from Rachel et al. (81), with permission.

circular genome that is heavily modified and can be cleaved by the methylation-dependent restriction endonuclease *Dpn*I, but not by the methylation-sensitive restriction endonuclease *Mbo*I (4). Therefore it appears that the viral genome, like the *Natrialba magadii* virus ΦCh1, encodes a *dam*-like methylase. There are no sequence data available for SNDV. The formation of SNDV is induced in the late stationary phase and a plaque test has not been established.

“STIV” (Unclassified)

Another novel virus-like particle was produced by a *Sulfolobus* isolate from YNP (88). Unlike all other known viruses of Crenarchaea, it has an icosahedral symmetry with unusual projections at its 5-fold axes of symmetry (figure 31-3). A 1 nm resolution structure for this virus has been determined by cryo-electron microscopy (88b) . The 72 nm diameter icosahedron has T=31 quasi-symmetry (18). The genome is double-stranded DNA about 19 kbp in size and it has been completely sequenced. None of the ORFs analyzed to date have any known homologs (G. Rice and M. Young, personal communication). The virus infects *S. solfataricus* strain P2 but its complete host range has not been determined. The proposed name is STIV (*Sulfolobus* turreted icosahedral virus). Similar icosahedral virus-like particles have been found in *Sulfolobus* enrichment cultures of samples from Lassen Volcanic National Park, California, USA (R. Diessner and K.M.S. unpublished data).

Uncultivated Viruses of Crenarchaeaota

Virus-like particles were detected in enrichment cultures from water samples from hot springs in YNP (81, 88). Some of these were shown to be produced by isolates belonging to the hyperthermophilic archaeal genus *Acidianus* (81). Some of these were morphologically similar to fuselloviruses, rudiviruses, and lipothrixviruses. Icosahedral particles similar to STIV were also observed. Additionally, particles with novel morphotypes were observed. The latter included 400 nm oval particles with long (up to 200 nm) projections at either end (figure 31-4G, H), small (32 nm) icosahedral particles, zipper-like particles built up of triangular subunits (figure 31-4D), and pleiomorphic particles with arrowlike heads and apparently helical tails (figure 31-4E, F). Rod-shaped helical particles with a central cavity (figure 31-4A) presumably belonged to *Rudiviridae*. Some types of filamentous particles, presumably belonging to the *Lipothrixviridae*, had unusual bulbous ends (figure 31-4B) or rounded tips (figure 31-4C) (81). Different virus-like particles were observed in PEG 6000 precipitates of cell-free filtrates of environmental samples from various Icelandic solfataric fields, some with spindle shapes, resembling SSVs, and others arrow-shaped, rod-shaped, and filamentous (129). In high-temperature neutral samples from YNP a number of virus-like particles have been observed which were often attached to *Thermofilum* and *Thermoproteus*-like organisms (W. Z., unpublished data).

Proviruses in Sequenced Archaeal Genomes

In addition to the almost complete copy of ΨM100 found in *Methanothermobacter wolfeii* as discussed above, a number of partial virus genomes have been found in the genomes of sequenced Archaea. The first to be found was a defective copy of the SSV1 genome in the genome of *S. shibatae* B12 (83), the strain from which SSV1 was originally isolated. By searching for disrupted integrase genes inserted into tRNA genes in a manner similar to SSV1, putative proviruses were found in the *Methanococcus jannashi* and *Pyrococcus horikoshii* genomes but not in those of *P. abyssi*, *P. furiosus*, *Archaeoglobus fulgidus*, or *Methanothermobacter thermoautotrophicum* (55). In the *S. solfataricus* P2 genome two putative integrated plasmids were found that contained disrupted integrase genes and genes from cryptic or conjugative plasmids (74). More potential integrated elements were found in the *Aeropyrum pernix*, *Archaeoglobus fulgidus*, *Halobacterium salinarum*, *S. solfataricus*, *S. acidocaldarius*, *Thermoplasma acidophilum*, and *Thermoplasma volcanium* genomes (96; reviewed in 95). The prevalence of these elements was proposed as a major means of horizontal gene transfer (96). None of these integrative elements, however, contained a complete prophage genome. Presumably the integrative genetic elements will only be mobile in the presence of an uninterrupted viral integrase. A partial integrated SSV genome, lacking an integrase gene, was found in the *S. solfataricus* genome by comparison of the SSV2 sequence to the genome (103). Apparently it has been trapped by the action of IS elements that are known to be very active in *Sulfolobus* (97).

Some Thoughts on the Origins and the Evolution of Viruses

Except within virus families or superfamilies (e.g., *Rudiviridae* + *Lipothrixviridae*) and in special cases of sharing mobile genes, and integrated proviruses, different viruses do not show sequence homology to each other or to other organisms. The widespread hypothesis that viruses are derived from cells is therefore not substantiated by close relationships of viral and organismal genes. Although this might in part be due to the higher rate of viral evolution, even in cases of significant homology of viral and organismal genes the viral homologs in phylogenetic dendrograms are often the lowest branches. This is particularly obvious for DNA metabolizing enzymes (27). Some viral proteins, such as the bacteriophage T3 and T7 DNA-dependent RNA polymerases and viral coat proteins, represent special solutions of evolutionary problems without homology in organisms. This deep branching of genes or complete lack of homology is to be expected when viruses or viral modules are not derived from organisms but are rather of prebiotic origins. The host for these viruses was the primordial soup or whatever place life first evolved. This is not as improbable as usually assumed, because the assembly of the first ancestor of all organisms also required the prior existence of all essential components in the prebiotic environment. Viruses or viral modules could then have developed from prebiotic macromolecules that were not integrated into the first common organismal ancestor but switched to organismal hosts when the “nutrients” of the primordial soup had been consumed by the expansion of organismal life.

Whereas all known organisms are clearly derived from one common ancestor, the conspicuous differences in the features of different virus families along with the lack of indications of close relationships between them suggest that viruses might be polyphyletic, well in accord with the assumption of prebiotic origins. This would imply their ancient ancestry.

A strong indication of an early origin of the archaeal and the bacterial *Myoviridae* is the close relationship of the halobacterial virus ΦH to *E. coli* phage P1, also belonging to the *Myoviridae*, suggesting a common origin rather than convergent evolution (91). It is improbable that the ancestors of the archaeal haloviruses or of the bacterial phages could have surmounted the strong and complex host-range barriers between their archaeal and bacterial hosts, respectively. More probably ancestral myoviruses were already infectious for ancestral organisms before the separation of the archaeal and the bacterial lineages and their progeny coevolved with the separating lineages of Archaea and Bacteria. Similarly, in the *Siphoviridae* family there is a clear phylogenetic relationship between some of the phages of dairy bacteria and the methanophage ΨM1. This relationship is apparent at the sequence level as well in the genomic organization (15).

Some of the similarities indicating common origins of distant virus families have not been recognized at the sequence level but are suggested by shared special features of virus life-styles. This has already been discussed above in comparing genome structure and replication of the archaeal rudiviruses and some large eukcaryal DNA viruses including poxviruses, African swine fever virus, and *Chlorella* viruses. Certain features of modules of such viruses indicate their even earlier appearance. Although the virion structure of fusellovirus SSV1 differs from that of myoviruses, its genome organization and life cycle resembles those of bacteriophage 186 or similar temperate bacteriophages belonging to the myoviruses. Gene expression is organized in two modules in a manner resembling early and late regulation in these phages. One of these, encompassing 16 "late" or viral protein synthesis and assembly genes, contains only one ORF with a cysteine codon while the other, encompassing genes involved in DNA replication and lysogeny, contains cysteine codons in 12 of 18 ORFs. Cysteine codons appeared supposedly late in the evolution of the genetic code (122), suggesting that the module practically devoid of them has an even earlier origin than the other, possibly well before the separation of the domains of life.

The major difficulty in establishing the very probable polyphyletic evolution of viruses, or their exchangeable modules respectively, appears to be the high rate of their evolution. Resolution of this difficulty will probably require the development of novel sophisticated methods of phylogenetic analysis.

References

1. Arnold, H. P. 1998. Isolierung und Beschreibung neuer Viren der crearchaealen Gattung *Sulfolobus*. PhD thesis, Ludwig Maximillians University, Munich, Germany.
2. Arnold, H. P., Q. She, H. Phan, K. Stedman, D. Prangishvili, I. Holz, J. K. Kristjansson, R. A. Garrett, and W. Zillig. 1999. The genetic element pSSVx of the extremely thermophilic crenarchaeon *Sulfolobus* is a hybrid between a plasmid and a virus. Mol. Microbiol. 34:217–226.
3. Arnold, H. P., K. M. Stedman, and W. Zillig. 1999. Archaeal phages, pp. 76–89. *In* A. Granoff and R. G. Webster (eds.) Encyclopedia of Virology, 2nd ed, vol. 1. Academic Press, London.
4. Arnold, H. P., U. Ziese, and W. Zillig. 2000. SNDV, a novel virus of the extremely thermophilic and acidophilic archaeon *Sulfolobus*. Virology 272:409–416.
5. Arnold, H. P., W. Zillig, U. Ziese, I. Holz, M. Crosby, T. Utterback, J. F. Weidmann, J. K. Kristjanson, H. P. Klenk, K. E. Nelson, and C. M. Fraser. 2000. A novel lipothrixvirus, SIFV, of the extremely thermophilic crenarchaeon *Sulfolobus*. Virology 267:252–266.
6. Balch, W. E., G. E. Fox, L. J. Magrum, C. R. Woese, and R. S. Wolfe. 1979. Methanogens: reevaluations of a unique biological group. Microbiol. Rev. 43:260–296.
7. Baranyi, U., R. Klein, W. Lubitz, D. H. Kruger, and A. Witte. 2000. The archaeal halophilic virus-encoded Dam-like methyltransferase M. ΦCh1-I methylates adenine residues and complements *dam* mutants in the low salt environment of *Escherichia coli*. Mol. Microbiol. 35:1168–1179.
8. Barns, S. M., C. F. Delwiche, J. D. Palmer, and N. R. Pace. 1996. Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. Proc. Natl. Acad. Sci. USA 93:9188–9193.
9. Bath, C., and M. L. Dyall-Smith. 1998. His1, an archaeal virus of the Fuselloviridae family that infects *Haloarcula hispanica*. J. Virol. 72:9392–9395.
10. Bertani, G. 1999. Transduction-like gene transfer in the methanogen *Methanococcus voltae*. J. Bacteriol. 181:2992–3002.
11. Bertani, G., and L. Baresi. 1986. Looking for gene transfer mechanisms in methanogenic bacteria, pp. 398. *In* O. Kandler and W. Zillig (eds.) Archaebacteria '85, vol. 1. Gustav Fischer, Stuttgart.
12. Bettstetter, M., X. Peng, R. A. Garrett, and D. Prangishvili. 2003. AFV1, a novel virus infecting hyperthermophilic archaea of the genus *Acidianus*. Virology 315:68–79.
13. Birkenbihl, R. P., K. Neef, D. Prangishvili, and B. Kemper. 2001. Holliday junction resolving enzymes of archaeal viruses SIRV1 and SIRV2. J. Mol. Biol. 309: 1067–1076.
14. Blum, H., W. Zillig, S. Mallok, H. Domdey, and D. Prangishvili. 2001. The genome of the archaeal virus SIRV1 has features in common with genomes of eukaryal viruses. Virology 281:6–9.
15. Brussow, H., and F. Desiere. 2001. Comparative phage genomics and the evolution of *Siphoviridae*: insights from dairy phages. Mol. Microbiol. 39:213–322.
16. Brussow, H., and R. W. Hendrix. 2002. Phage genomics: small is beautiful. Cell 108:13–16.
17. Cannio, R., P. Contursi, M. Rossi, and S. Bartolucci. 1998. An autonomously replicating transforming vector for *Sulfolobus solfataricus*. J. Bacteriol. 180:3237–3240.
18. Caspar, D. L. D., and A. Klug. 1962. Physical principles in the construction of regular viruses. Cold Spring Harb. Symp. Quant. Biol. 27:1–24.
19. Daniels, L. L., and A. C. Wais. 1990. Ecophysiology of bacteriophage s5100 infecting *Halobacterum cutirubrum*. Appl. Environ. Microbiol. 56:3605–3608.
20. Daniels, L. L., and A. C. Wais. 1984. Restriction and modification of halophage S45 in *Halobacterium*. Curr. Microbiol. 10:133–136.
21. Daniels, L. L., and A. C. Wais. 1998. Virulence in phage populations infecting *Halobacterium cutirubrum*. FEMS Microb. Ecol. 25:129–134.
22. DeLong, E. F. 1998. Everything in moderation: archaea as "non-extremophiles". Curr. Opin. Genet. Dev. 8:649–654.
23. Dyall-Smith, M., S. L. Tang, and C. Bath. 2003. Haloarchaeal viruses: How diverse are they? Res. Microbiol. 154:309–313.
24. Eiserling, F., A. Pushkin, M. Gingery, and G. Bertani. 1999. Bacteriophage-like particles associated with the gene transfer agent of *Methanococcus voltae* PS. J. Gen. Virol. 80:3305–3308.
25. Esposito, D., W. P. Fitzmaurice, R. C. Benjamin, S. D. Goodman, A. S. Waldman, and J. J. Scocca. 1996. The complete nucleotide sequence of bacteriophage HP1 DNA. Nucleic Acids Res. 24:2360–2368.
26. Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, J. M. Merrick, K. McKenney, G. Sutton, W. FitzHugh, C. Fields, J. D. Gocayne, J. Scott, R. Shirley, L.-I. Liu, A. Glodek, J. M. Kelley, J. F. Weidman, C. A. Phillips, T. Spriggs, E. Hedblom, M. D. Cotton, T. R. Utterback, M. C. Hanna, D. T. Nguyen, D. M. Saudek, R. C. Brandon, L. D. Fine, J. L. Fritchman, J. L. Fuhrmann, N. S. M. Geoghagen, C. L. Gnehm, L. A. McDonald, K. V. Small, C. M. Fraser,

H. O. Smith, and J. C. Venter. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269:496–512.

27. Forterre, P. 1999. Displacement of cellular proteins by functional analogues from plasmids or viruses could explain puzzling phylogenies of many DNA informational proteins. Mol. Microbiol. 33:457–465.
28. Garrity, G. M., and J. G. Holt. 2001. Phylum AI. Crenarchaeota *phy. nov.*, p. 169. *In* D. Boone, R. Castenholz, and G. Garrity (eds.) Bergey's Manual of Systematic Bacteriology, 2nd edn, vol. 1. Springer, New York.
29. Garrity, G. M., and J. G. Holt. 2001. Phylum AII. Euryarchaeota *phy. nov.*, pp. 169. *In* D. Boone, R. Castenholz, and G. Garrity (eds.) Bergey's Manual of Systematic Bacteriology, 2nd edn, vol. 1. Springer, New York.
30. Gelderblom, H. R. 1991. Assembly and morphology of HIV: potential effect of structure on viral function. AIDS 5:617–637.

30a. Geslin, C., M. Le Romancer, M. Gaillard, G. Erauso, and D. Prieur. 2003. Observation of virus-like particles in high temperature enrichment cultures from deep-sea hydrothermal vents. Res. Microbiol. 154:303–307.

30b. Geslin, C., M. Le Romancer, G. Erauso, M. Gaillard, G. Perrot, and D. Prieur. 2003. PAV1, the first virus-like particle isolated from a hyperthermophilic euryarchaeote, "*Pyrococcus abyssi*." J. Bacteriol. 185:3888–3894.

31. Grant, W. D., M. Kamekura, T. J. McGenity, and A. Ventosa. 2001. Class III. Halobacteria *class nov.*, pp. 294–334. *In* D. Boone, R. Castenholz, and G. Garrity (eds.) Bergey's Manual of Systematic Bacteriology, 2nd edn, vol. 1. Springer, New York.
32. Grogan, D. W. 1989. Phenotypic characterization of the archaebacterial genus *Sulfolobus*: comparison of five wild-type strains. J. Bacteriol. 171:6710–6719.
33. Gropp, F., B. Grampp, P. Stolt, P. Palm, and W. Zillig. 1992. The immunity-conferring plasmid p Φ HL from the *Halobacterium salinarium* phage ΦH: nucleotide sequence and transcription. Virology 190:45–54.
34. Guixa-Boixareu, N., J. I. Calderón-Paz, M. Heldal, G. Bratbak, and C. Pedrós-Alió. 1996. Viral lysis and bacterivory as prokaryotic loss factors along a salinity gradient. Aquat. Microb. Ecol. 11:215–227.
35. Gurevich, P., and A. Oren. 1993. Characterization of the dominant halophilic archaea in a bacterial bloom in the Dead Sea. FEMS Microbial Ecol. 12:249–256.
36. Hatfull, G. F., and G. J. Sarkis. 1993. DNA sequence, structure and gene expression of mycobacteriophage L5: a phage system for mycobacterial genetics. Mol. Microbiol. 7:395–405.
37. Hendrix, R. W. 2002. Bacteriophages: evolution of the majority. Theor. Popul. Biol. 61:471–480.
38. Huber, H., M. J. Hohn, R. Rachel, T. Fuchs, V. C. Wimmer, and K. O. Stetter. 2002. A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. Nature 417:63–67.
39. Huber, H., and K. O. Stetter. 2001. Family *Sulfolobaceae*, pp. 198. *In* D. Boone, R. Castenholz, and G. Garrity (eds.) Bergey's Manual of Systematic Bacteriology, 2nd edn, vol. 1. Springer, New York.
40. Ishino, Y., and S. Ishino. 2001. DNA polymerases from euryarchaeota. Methods Enzymol. 334:249–260.
41. Iwai, T., N. Kurosawa, Y. H. Itoh, and T. Horiuchi. 2000. Phylogenetic analysis of archaeal PCNA homologues. Extremophiles 4:357–364.
42. Janekovic, D., S. Wunderl, I. Holz, W. Zillig, A. Gierl, and H. Neumann. 1983. TTV-1, TTV-2 and TTV-3: a family of viruses of the extremely thermophilic anaerobic sulfur reducing archaebacterium *Thermoproteus tenax*. Mol. Gen. Genet. 192:39–45.
43. Jonuscheit, M., E. Martusewitsch, K. M. Stedman, and C. Schleper. 2003. A reporter gene system for the hyperthermophilic archaeon *Sulfolobus solfataricus* based on a selectable and integrative shuttle vector. Mol. Microbiol. 48:1241–1252.
44. Jordan, M., L. Meile, and T. Leisinger. 1989. Organization of *Methanobacterium thermoautotrophicum* bacteriophage ΨM1 DNA. Mol. Gen. Genet. 220:161–164.
45. Karner, M. B., E. F. DeLong, and D. M. Karl. 2001. Archaeal dominance in the mesopelagic zone of the Pacific Ocean. Nature 409:507–510.
46. Keeling, P. J., H. P. Klenk, R. K. Singh, O. Feeley, C. Schleper, W. Zillig, W. F. Doolittle, and C. W. Sensen. 1996. Complete nucleotide sequence of the *Sulfolobus islandicus* multicopy plasmid pRN1. Plasmid 35:141–144.
47. Keeling, P. J., H. P. Klenk, R. K. Singh, M. E. Schenk, C. W. Sensen, W. Zillig, and W. F. Doolittle. 1998. *Sulfolobus islandicus* plasmids pRN1 and pRN2 share distant but common evolutionary ancestry. Extremophiles 2:391–393.
48. Klein, R., U. Baranyi, N. Rossler, B. Greineder, H. Scholz, and A. Witte. 2002. *Natrialba magadii* virus ΦCh1: first complete nucleotide sequence and functional organization of a virus infecting a haloalkaliphilic archaeon. Mol. Microbiol. 45:851–863.
49. Klein, R., B. Greineder, U. Baranyi, and A. Witte. 2000. The structural protein E of the archaeal virus phiCh1: evidence for processing in *Natrialba magadii* during virus maturation. Virology 276:376–387.
50. Knox, M. R., and J. E. Harris. 1986. Isolation and characterization of a bacteriophage of *Methanobrevibacter smithii*. *In* Abstracts of the XIV International Congress on Microbiology. IUMS, Manchester.
51. Lawrence, J. G., G. F. Hatfull, and R. W. Hendrix. 2002. Imbroglios of viral taxonomy: genetic exchange and failings of phenetic approaches. J. Bacteriol. 184: 4891–4905.
52. Lucas, S., L. Toffin, Y. Zivanovic, D. Charlier, H. Moussard, P. Forterre, D. Prieur, and G. Erauso. 2002. Construction of a shuttle vector for, and spheroplast transformation of, the hyperthermophilic archaeon *Pyrococcus abyssi*. Appl. Environ. Microbiol. 68:5528–5536.
53. Luo, Y., P. Pfister, T. Leisinger, and A. Wasserfallen. 2001. The genome of archaeal prophage ΨM100 encodes the lytic enzyme responsible for autolysis of *Methanothermobacter wolfeii*. J. Bacteriol. 183:5788–5792.
54. Luo, Y., P. Pfister, T. Leisinger, and A. Wasserfallen. 2002. Pseudomurein endoisopeptidases PeiW and PeiP, two moderately related members of a novel family of proteases

produced in *Methanothermobacter* strains. FEMS Microbiol. Lett. 208:47–51.

55. Makino, S., N. Amano, H. Koike, and M. Suzuki. 1999. Prophages inserted in archaebacterial genomes. Proc. Jpn. Acad., Ser. B 75:166–171.
56. Marteinsson, V. T., J. K. Kristjansson, H. Kristmannsdottir, M. Dahlkvist, K. Saemundsson, M. Hannington, S. K. Petursdottir, A. Geptner, and P. Stoffers. 2001. Discovery and description of giant submarine smectite cones on the seafloor in Eyjafjordur, northern Iceland, and a novel thermal microbial habitat. Appl. Environ. Microbiol. 67:827–833.
57. Martin, A., S. Yeats, D. Janekovic, W. D. Reiter, W. Aicher, and W. Zillig. 1984. SAV-1, a temperate UV inducible DNA virus-like particle from the archaebacterium *Sulfolobus acidocaldarius* isolate B-12. EMBO J. 3:2165–2168.
58. Meile, L., P. Abendschein, and T. Leisinger. 1990. Transduction in the archaebacterium *Methanobacterium thermoautotrophicum* Marburg. J. Bacteriol. 172:3507–3508.
59. Meile, L., U. Jenal, D. Studer, M. Jordan, and T. Leisinger. 1989. Characterization of ΨM1, a virulent phage of *Methanobacterium thermoautotrophicum* Marburg. Arch. Microbiol. 152:105–110.
60. Muskhelishvili, G. 1994. The archaeal SSV integrase promotes intermolecular excisive recombination in vitro. Appl. Microbiol. 16:605–608.
61. Muskhelishvili, G., P. Palm, and W. Zillig. 1993. SSV1-encoded site-specific recombination system in *Sulfolobus shibatae*. Mol. Gen. Genet. 237:334–342.
62. Nadal, M., G. Mirambeau, P. Forterre, W. D. Reiter, and M. Duguet. 1986. Positively supercoiled DNA in a virus-like particle of an archaebacterium. Nature 321: 256–258.
63. Neumann, H., and W. Zillig. 1990. Structural variability in the genome of the *Thermoproteus tenax* virus TTV1. Mol. Gen. Genet. 222:435–437.
64. Ng, W. V., S. P. Kennedy, G. G. Mahairas, B. Berquist, M. Pan, H. D. Shukla, S. R. Lasky, N. S. Baliga, V. Thorsson, J. Sbrogna, S. Swartzell, D. Weir, J. Hall, T. A. Dahl, R. Welti, Y. A. Goo, B. Leithauser, K. Keller, R. Cruz, M. J. Danson, D. W. Hough, D. G. Maddocks, P. E. Jablonski, M. P. Krebs, C. M. Angevine, H. Dale, T. A. Isenbarger, R. F. Peck, M. Pohlschroder, J. L. Spudich, K. W. Jung, M. Alam, T. Freitas, S. Hou, C. J. Daniels, P. P. Dennis, A. D. Omer, H. Ebhardt, T. M. Lowe, P. Liang, M. Riley, L. Hood, and S. DasSarma. 2000. Genome sequence of *Halobacterium* species NRC-1. Proc. Natl. Acad. Sci. USA 97:12176–12181.
65. Nolling, J., and W. M. de Vos. 1992. Identification of the CTAG-recognizing restriction-modification systems MthZI and MthFI from *Methanobacterium thermoformicicum* and characterization of the plasmid-encoded mthZIM gene. Nucleic Acids Res. 20:5047–5052.
66. Nolling, J., A. Groffen, and W. M. De Vos. 1993. ΦF1 and ΦF3, two novel virulent, archaeal phages infecting different thermophilic strains of the genus *Methanobacterium*. J. Gen. Microbiol. 139:2511–2516.
67. Nuttall, S. D., and M. L. Dyall-Smith. 1995. Halophage HF2: genome organization and replication strategy. J. Virol. 69:2322–2327.
68. Nuttall, S. D., and M. L. Dyall-Smith. 1993. HF1 and HF2: novel bacteriophages of halophilic archaea. Virology 197:678–684.
69. Oren, A. 1994. The ecology of the extremely halophilic archaea. FEMS Microbiol. Rev. 13:415–440.
70. Oren, A., G. Bratbak, and M. Heldal. 1997. Occurrence of virus-like particles in the Dead Sea. Extremophiles 1:143–149.
71. Palm, P., C. Schleper, B. Grampp, S. Yeats, P. McWilliam, W. D. Reiter, and W. Zillig. 1991. Complete nucleotide sequence of the virus SSV1 of the archaebacterium *Sulfolobus shibatae*. Virology 185:242–250.
72. Pauling, C. 1982. Bacteriophages of *Halobacterium halobium*: isolated from fermented fish sauce and primary characterization. Can. J. Microbiol. 28:916–921.
73. Peng, X., H. Blum, Q. She, S. Mallok, K. Brugger, R. A. Garrett, W. Zillig, and D. Prangishvili. 2001. Sequences and replication of genomes of the archaeal rudiviruses SIRV1 and SIRV2: relationships to the archaeal lipothrixvirus SIFV and some eukaryal viruses. Virology 291:226–234.
74. Peng, X., I. Holz, W. Zillig, R. A. Garrett, and Q. She. 2000. Evolution of the family of pRN plasmids and their integrase-mediated insertion into the chromosome of the crenarchaeon *Sulfolobus solfataricus*. J. Mol. Biol. 303:449–454.
75. Pfeifer, F. 1987. Genetics of Halobacteria, pp. 105–133. *In* F. Rodriguez-Valera (ed.) Halophilic Bacteria. CRC Press, Boca Raton, Fla.
76. Pfister, P., A. Wasserfallen, R. Stettler, and T. Leisinger. 1998. Molecular analysis of *Methanobacterium* phage ΨM2. Mol. Microbiol. 30:233–244.
77. Prangishvili, D., H. P. Arnold, D. Gotz, U. Ziese, I. Holz, J. K. Kristjansson, and W. Zillig. 1999. A novel virus family, the Rudiviridae: structure, virus–host interactions and genome variability of the sulfolobus viruses SIRV1 and SIRV2. Genetics 152:1387–1396.
78. Prangishvili, D., H. P. Klenk, G. Jakobs, A. Schmiechen, C. Hanselmann, I. Holz, and W. Zillig. 1998. Biochemical and phylogenetic characterization of the dUTPase from the archaeal virus SIRV. J. Biol. Chem. 273:6024–6029.
79. Prangishvili, D., K. Stedman, and W. Zillig. 2001. Viruses of the extremely thermophilic archaeon *Sulfolobus*. Trends Microbiol. 9:39–43.
80. Qureshi, S. A., and S. P. Jackson. 1998. Sequence-specific DNA binding by the *S. shibatae* TFIIB homolog, TFB, and its effect on promoter strength. Mol. Cell 1:389–400.
81. Rachel, R., M. Bettstetter, B. P. Hedlund, M. Haring, A. Kessler, K. O. Stetter, and D. Prangishvili. 2002. Remarkable morphological diversity of viruses and virus-like particles in hot terrestrial environments. Arch. Virol. 147: 2419–2429.
82. Reiter, W. D., U. Hudepohl, and W. Zillig. 1990. Mutational analysis of an archaebacterial promoter: essential role of a TATA box for transcription efficiency and start-site selection in vitro. Proc. Natl. Acad. Sci. USA 87:9509–9513.
83. Reiter, W. D., and P. Palm. 1990. Identification and characterization of a defective SSV1 genome integrated into a tRNA gene in the archaebacterium *Sulfolobus* sp. B12. Mol. Gen. Genet. 221:65–71.

84. Reiter, W. D., P. Palm, A. Henschen, F. Lottspeich, W. Zillig, and B. Grampp. 1987. Identification and characterization of the genes encoding three structural proteins of the *Sulfolobus* virus-like particle SSV1. Mol. Gen. Genet. 206:144–153.

85. Reiter, W. D., P. Palm, and S. Yeats. 1989. Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. Nucleic Acids Res. 17:1907–1914.

86. Reiter, W. D., P. Palm, S. Yeats, and W. Zillig. 1987. Gene expression in archaebacteria physical mapping of constitutive and UV-inducible transcripts from the *Sulfolobus* virus-like particle SSV1. Mol. Gen. Genet. 209:270–275.

87. Reysenbach, A. L., M. Ehringer, and K. Hershberger. 2000. Microbial diversity at 83 °C in Calcite Springs, Yellowstone National Park: another environment where the *Aquificales* and “Korarchaeota” coexist. Extremophiles 4:61–67.

88a. Rice, G., K. M. Stedman, J. Snyder, B. Wiedenheft, D. Willits, S. Brumfield, T. McDermott, and M. J. Young. 2001. Novel viruses from extreme thermal environments. Proc. Natl. Acad. Sci. USA 98:13341–13345.

88b. Rice, G., L. Tang, K. Stedman, F. Roberto, J. Spuhler, E. Gillitzer, J. E. Johnson, T. Douglas, and M. Young. 2004. The structure of a thermophilic archaeal virus shows a double-stranded DNA viral capsid type that spans all domains of life. Proc. Natl. Acad. Sci. USA 10.1:7716–7720.

89. Schleper, C., K. Kubo, and W. Zillig. 1992. The particle SSV1 from the extremely thermophilic archaeon *Sulfolobus* is a virus: demonstration of infectivity and of transfection with viral DNA. Proc. Natl. Acad. Sci. USA 89:7645–7649.

90. Schnabel, H. 1984. Integration of plasmid p-ΦHl into phage genomes during infection of *Halobacterium halobium* R-1-L with phage ΦHl. Mol. Gen. Genet. 197: 19–23.

91. Schnabel, H., and W. Zillig. 1984. Circular structure of the genome of phage ΦH in a lysogenic *Halobacterium halobium*. Mol. Gen. Genet. 193:422–426.

92. Schnabel, H., and W. Zillig. 1982. Structural variations in the DNA of *Halobacterium halobium* phage ΦH. Zentralbl. Bakteriol. Mikrobiol. Hyg., 1, Abt. Orig., A. 253:35–36.

93. Schnabel, H., W. Zillig, M. Pfaeffle, R. Schnabel, H. Michel, and H. Delius. 1982. *Halobacterium halobium* phage ΦH. EMBO J. 1:87–92.

94. Segerer, A., K. O. Stetter, and F. Klink. 1985. Two contrary modes of chemolithotrophy in the same archaebacterium. Nature 313:787–789.

95. She, Q., K. Brugger, and L. Chen. 2002. Archaeal integrative genetic elements and their impact on genome evolution. Res. Microbiol. 153:325–332.

96. She, Q., X. Peng, W. Zillig, and R. A. Garrett. 2001. Gene capture in archaeal chromosomes. Nature 409:478.

97. She, Q., R. K. Singh, F. Confalonieri, Y. Zivanovic, G. Allard, M. J. Awayez, C. C.-Y. Chan-Weihere, I. G. Clausen, B. A. Curtis, A. De Moors, G. Erauso, C. Fletcher, P. M. K. Gordon, I. Heikamp-de Jong, A. C. Jeffries, C. J. Kozera, N. Medina, X. Peng, H. P. Thi-Ngoc, P. Redder, M. E. Schenk, C. Theriault, N. Tolstrup, R. L. Charlebois, W. F. Doolittle, M. Duguet, T. Gaasterland, R. A. Garrett, M. A. Ragan, C. W. Sensen, and J. Van der Oost. 2001. The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. Proc. Natl. Acad. Sci. USA 98:7835–7840.

98. She, Q. X., H. E. Phan, R. A. Garrett, S. V. Albers, K. M. Stedman, and W. Zillig. 1998. Genetic profile of pNOB8 from *Sulfolobus*: the first conjugative plasmid from an archaeon. Extremophiles 2:417–425.

99. Siebers, B., V. F. Wendisch, and R. Hensel. 1997. Carbohydrate metabolism in *Thermoproteus tenax*: in vivo utilization of the non-phosphorylative Entner–Doudoroff pathway and characterization of its first enzyme, glucose dehydrogenase. Arch. Microbiol. 168:120–127.

100. Smith, D. R., L. A. Doucette-Stamm, C. Deloughery, H. Lee, J. Dubois, T. Aldredge, R. Bashirzadeh, D. Blakely, R. Cook, K. Gilbert, D. Harrison, L. Hoang, P. Keagle, W. Lumm, B. Pothier, D. Qiu, R. Spadafora, R. Vicaire, Y. Wang, J. Wierzbowski, R. Gibson, N. Jiwani, A. Caruso, D. Bush, J. N. Reeve, et al. 1997. Complete genome sequence of *Methanobacterium thermoautotrophicum* ΔH: functional analysis and comparative genomics. J. Bacteriol. 179:7135–7155.

101. Stax, D., R. Hermann, R. Falchetto, and T. Leisinger. 1992. The lytic enzyme in bacteriophage ΨM1 induced lysates of *Methanobacterium thermoautotrophicum* Marburg. FEMS Microbial Lett. 100:433–438.

102. Stedman, K., C. Schleper, E. Rumpf, and W. Zillig. 1999. Genetic requirements for the function of the archaeal virus SSV1 in *Sulfolobus solfataricus*: construction and testing of viral shuttle vectors. Genetics 152:1397–1405.

103. Stedman, K. M., Q. She, H. Phan, H. P. Arnold, I. Holz, R. A. Garrett, and W. Zillig. 2003. Relationships between fuselloviruses infecting the extremely thermophilic archaeon *Sulfolobus*: SSV1 and SSV2. Res. Microbiol. 154:295–302.

104. Stettler, R., C. Thurner, D. Stax, L. Meile, and T. Leisinger. 1995. Evidence for a defective prophage on the chromosome of *Methanobacterium wolfei*. FEMS Microbiol. Lett. 132:85–89.

105. Stolt, P., B. Grampp, and W. Zillig. 1994. Genes for DNA cytosine methyltransferases and structural proteins, expressed during lytic growth by the phage ΦH of the archaebacterium *Halobacterium salinarium*. Biol. Chem. Hoppe Seyler 375:747–757.

106. Stolt, P., and W. Zillig. 1993. Antisense RNA mediates transcriptional processing in an archaebacterium, indicating a novel kind of RNase activity. Mol. Microbiol. 7:875–882.

107. Stolt, P., and W. Zillig. 1995. Archaebacterial bacteriophages. *In* R. Webster and A. Granoff (eds.) Encyclopedia of Virology Plus. Academic Press, London.

108. Stolt, P., and W. Zillig. 1994. Gene regulation in halophage ΦH; more than promoters. Appl. Microbiol. 16:591–596.

109. Stolt, P., and W. Zillig. 1994. Transcription of the halophage ΦH repressor gene is abolished by transcription from an inversely oriented lytic promoter. FEBS Lett. 344:125–128.

110. Tang, S. L., S. Nuttall, K. Ngui, C. Fisher, P. Lopez, and M. Dyall-Smith. 2002. HF2: a double-stranded DNA tailed haloarchaeal virus with a mosaic genome. Mol. Microbiol. 44:283–296.

111. Torsvik, T., and I. D. Dundas. 1974. Bacteriophage of *Halobacterium salinarum*. Nature 248:680–681.
112. Torsvik, T., and I. D. Dundas. 1980. Persisting phage infection in *Halobacterium salinarum* str.1. J. Gen. Virol. 47:29–36.
113. Touzel, J. P., E. C. De Macario, J. Nolling, W. M. De Vos, T. Zhilina, and A. M. Lysenko. 1992. DNA relatedness among some thermophilic members of the genus *Methanobacterium*: emendation of the species *Methanobacterium thermoautotrophicum* and rejection of *Methanobacterium thermoformicicum* as a synonym of *Methanobacterium thermoautotrophicum*. Int. J. Syst. Bacteriol. 42:408–411.
114. Tumbula, D. L., T. L. Bowen, and W. B. Whitman. 1995. Growth of methanogens on solidified medium, pp. 49–55. *In* K. R. Sowers and H. J. Schreier (eds.) Archaea: Methanogens. A Laboratory Manual. Cold Spring Harbor Laboratory Press, Plainview.
115. Tumbula, D. L., and W. B. Whitman. 1999. Genetics of *Methanococcus*: possibilities for functional genomics in Archaea. Mol. Microbiol. 33:1–7.
116. Vogelsang-Wenke, H., and D. Oesterhelt. 1986. Halophage ΦN, pp. 403–405. *In* O. Kandler and W. Zillig (eds.) Archaebacteria '85, vol. 1. Gustav Fischer, Stuttgart.
117. Wais, A. C., M. Kon, R. E. MacDonald, and B. D. Stollar. 1975. Salt-dependent bacteriophage infecting *Halobacterium cutirubrum* and *H. halobium*. Nature 256:314–315.
118. Wasserfallen, A., J. Nolling, P. Pfister, J. Reeve, and E. Conway de Macario. 2000. Phylogenetic analysis of 18 thermophilic *Methanobacterium* isolates supports the proposals to create a new genus, *Methanothermobacter* gen. nov., and to reclassify several isolates in three species, *Methanothermobacter thermautotrophicus* comb. nov., *Methanothermobacter wolfeii* comb. nov., and *Methanothermobacter marburgensis* sp. nov. Int. J. Syst. Evol. Microbiol. 50:43–53.
119. Wiedenheft, B., K. Stedman, F. Roberto, D. Willits, A. K. Gleske, L. Zoeller, J. Snyder, T. Douglas, and M. Young. 2004. Comparative genomic analysis of hyperthermophilic archaeal *Fuselloviridae* viruses. J. Virol. 78:1954–1961.
120. Witte, A., U. Baranyi, R. Klein, M. Sulzner, C. Luo, G. Wanner, D. H. Kruger, and W. Lubitz. 1997. Characterization of *Natronobacterium magadii* phage ΦCh1, a unique archaeal phage containing DNA and RNA. Mol. Microbiol. 23:603–616.
121. Woese, C. R., and G. E. Fox. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc. Natl. Acad. Sci. USA 74:5088–5090.
122. Wong, J. T. 1975. A co-evolution theory of the genetic code. Proc. Natl. Acad. Sci. USA 72:1909–1912.
123. Wood, A. G., W. B. Whitman, and J. Konisky. 1989. Isolation and characterization of an archaebacterial viruslike particle from *Methanococcus voltae* A3. J. Bacteriol. 171:93–98.
124. Yeats, S., P. McWilliam, and W. Zillig. 1982. A plasmid in the archaebacterium *Sulfolobus solfataricus*. EMBO J. 1:1035–1038.
125. Zillig, W. 1991. Comparative biochemistry of Archaea and Bacteria. Curr. Opin. Genet. Dev. 1:544–551.
126. Zillig, W., H. P. Arnold, I. Holz, D. Prangishvili, A. Schweier, K. Stedman, Q. She, H. Phan, R. Garrett, and J. K. Kristjansson. 1998. Genetic elements in the extremely thermophilic archaeon *Sulfolobus*. Extremophiles 2: 131–140.
127. Zillig, W., F. Gropp, A. Henschen, H. Neumann, P. Palm, W. D. Reiter, M. Rettenberger, H. Schnabel, and S. Yeats. 1985. Archaebacterial virus host systems. Appl. Microbiol. 7:58–66.
128. Zillig, W., I. Holz, H. P. Klenk, J. Trent, S. Wunderl, D. Janekovic, E. Imsel, and B. Haas. 1987. *Pyrococcus woesei* new-species an ultra-thermophilic marine archaebacterium representing a novel order *Thermococcales*. Appl. Microbiol. 9:62–70.
129. Zillig, W., A. Kletzin, C. Schleper, I. Holz, D. Janekovic, J. Hain, M. Lanzendoerfer, and J. K. Kristjansson. 1994. Screening for *Sulfolobales*, their plasmids and their viruses in Icelandic solfataras. Appl. Microbiol. 16: 609–628.
130. Zillig, W., D. Prangishvilli, C. Schleper, M. Elferink, I. Holz, S. Albers, D. Janekovic, and D. Götz. 1996. Viruses, plasmids and other genetic elements of thermophilic and hyperthermophilic Archaea. FEMS Microbiol. Rev. 18:225–236.
131. Zillig, W., W. D. Reiter, P. Palm, F. Gropp, H. Neumann, and M. Rettenberger. 1988. Viruses of Archaebacteria, pp. 517–558. *In* R. Calendar (ed.) The Bacteriophages, vol. 1. Plenum Press, New York.
132. Zillig, W., and A. L. Reysenbach. 2001. Thermococcales, pp. 341. *In* D. Boone, R. Castenholz, and G. Garrity (eds.) Bergey's Manual of Systematic Bacteriology, 2nd edn, vol. 1. Springer, New York.
133. Zillig, W., J. Tu, and I. Holz. 1981. *Thermoproteales*: a third order of thermoacidophilic archaebacteria. Nature 293:85–86.
134. Zillig, W., S. Yeats, I. Holz, A. Boeck, M. Rettenberger, F. Gropp, and G. Simon. 1986. *Desulfurolobus ambivalens* new-genus new-species an autotrophic archaebacterium facultatively oxidizing or reducing sulfur. Appl. Microbiol. 8:197–203.
135. Zivanovic, Y., P. Lopez, H. Philippe, and P. Forterre. 2002. *Pyrococcus* genome comparison evidences chromosome shuffling-driven evolution. Nucleic Acids Res. 30:1902–1910.

32

Phages of Cyanobacteria

NICHOLAS H. MANN

The scientific importance of the phages of cyanobacteria (cyanophages) is intimately associated with the ecological significance of their hosts. Cyanobacteria are arguably the most diverse and widely distributed group of eubacteria on the planet and play central roles in major biogeochemical processes, such as the carbon and nitrogen cycles. Cyanobacteria exist in a wide range of freshwater and marine environments, ranging from thermophilic to psycrophilic, and terrestrial environments, including those subject to periodic desiccation. By virtue of their higher plant-like oxygenic photosynthetic apparatus they contribute significantly to the maintenance of the Earth's atmosphere, in terms of both oxygen production and carbon dioxide fixation. Consequently, the ability of cyanophages to determine the population structures and genetic diversity of cyanobacteria, as well as to potentially influence the dynamics of biogeochemical processes, gives them a unique ecological significance.

The study of cyanobacteria has a long history. Indeed, their taxonomy was discussed by Linnaeus in 1753 (61; as cited by 124). Thus, it seems strange that the first cyanophage, which caused lysis of several species of cyanobacteria of the genera *Lyngbya, Plectonema,* and *Phormidium,* was not discovered until 1963 (by Safferman and Morris (90), during the characterization of sewage settling ponds in the neighborhood of Cincinnati for the Environmental Protection Agency). Why were phages infecting cyanobacteria not found prior to 1963? One possible reason is that before 1963 cyanobacteria were called blue-green algae and considered to be eukaryotic. Subsequently the cyanobacteria were shown to have lysozyme-sensitive cell walls, to be sensitive to penicillin, to have diaminopimelic acid in their peptidoglycan, and to have typical prokaryotic ribosomes. These features established their prokaryotic nature. The discovery of phages that could be characterized by typical one-step growth curves and counted by the formation of plaques on confluent lawns of the host helped to confirm this assignment.

Following the recognition that viruses were abundant in sea water (12) and that marine cyanobacteria were visibly phage-infected (82), the first marine cyanophages were isolated in 1991 (72) while the first phages infecting the abundant marine *Synechococcus* strains were isolated in 1993 (112, 121, 128). There is an enormous diversity of cyanobacteria in the marine environment, which is well reviewed by Golubic et al. (43). Work on marine cyanophages, however, has been focused largely, though not exclusively, on those which infect unicellular cyanobacteria of the genera *Synechococcus* and *Prochlorococcus*. These organisms dominate the prokaryotic component of the picophytoplankton and make a very significant contribution to overall primary production in the oceans (40, 60, 62, 118). Cyanophages in general and marine cyanophages in particular have been the topics of recent reviews (66, 110). In this chapter there is no attempt to comprehensively review the cyanophage literature, but rather attention is focused on the present state of knowledge regarding the nature of cyanophages, their interactions with their host, and their ecological impact.

The Cyanobacterial Hosts

For a variety of reasons an account of cyanophages must begin with a consideration of the hosts they infect. The cyanobacteria occupy an extraordinarily diverse range of environments and often make major contributions to biogeochemical processes. Their oxygenic photoautotrophic mode of nutrition is distinct in the prokaryotic kingdom, being shared only with the prochlorophytes, and has significant implications for phage replication. Their cell envelopes have certain structural features which distinguish them from other eubacteria and which influence phage adsorption. Finally, many areas of the taxonomy of cyanobacteria are highly problematic and this has led to confusion in some host range studies.

The most important physiological feature of cyanobacteria is oxygenic photosynthesis. Like higher plants, cyanobacteria possess two photosystems, PSI and PSII, that are connected by an inter-system electron transport pathway. Of the cyanobacteria that have been tested, the very large majority are capable of taking up and utilizing exogenous carbon compounds, but very few are capable of dark growth at the expense of exogenous carbon compounds. Thus, cyanobacteria, for the most part, are obligate phototrophs.

A second characteristic feature of many cyanobacteria is the possession of macromolecular light-harvesting antennae, known as phycobilisomes, that are composed of chromophore-bearing phycobiliproteins. It is these phycobiliproteins, together with chlorophyll *a*, that give cyanobacteria their characteristic coloration; blue-green when phycocyanin is the major phycobiliprotein and orange-red when phycoerythrin predominates. The phycobiliproteins are extremely abundant and may represent as much as half the total cell protein. A major exception is the lack of phycobilisomes in marine *Prochlorococcus* strains and their possession of a chlorophyll a_2/b_2 light-harvesting complex (23).

Obviously, in the natural environment cyanobacteria are subject to a light–dark cycle. During the dark periods they generate maintenance energy by the oxidation of carbohydrate reserves, usually glycogen, via the pentose phosphate pathway. Many species of cyanobacteria—including the ecologically important filamentous marine cyanobacterium *Trichodesmium*—are also capable of fixing atmospheric dinitrogen.

Cyanobacteria exhibit a range of patterns of cellular organization, from simple unicells that divide by binary fission through to multiseriate trichomes with branches and differentiated cells. These morphological features have been used extensively to produce a cyanobacterial taxonomy. However, the classification of cyanobacteria is extremely problematic and the factors leading to this are well reviewed by Wilmotte (125). For many years cyanobacteria, or blue-green algae as they were then known, were treated as just one group of the algae, with morphological characters being the key to their taxonomy and their nomenclature being ruled by the Botanical Code. This classical botanical taxonomy was challenged by the recognition of the distinction between prokaryotes and eukaryotes and the compelling evidence that blue-green algae were in fact bacteria and should be subject to the Bacteriological Code (108). Accordingly, a new bacteriological taxonomy of the cyanobacteria was published by Rippka et al. in 1979 (85).

Currently, cyanobacteria that have been studied in pure culture are placed in five orders (119), but molecular approaches show that three of these are not necessarily monophyletic (e.g., 50) and only two form single coherent lineages. The situation is further complicated by the fact that many ecologists still adhere to the Botanical Code when naming cyanobacteria in field samples (124). One highly relevant result of this taxonomic confusion is that in certain cases it has led to the erroneous interpretation of phage host range studies (110).

The marine unicellular cyanobacteria are a particularly important group as far as phage biology is concerned. Also, their comparatively recent discovery and simple morphology mean that there are no serious taxonomic problems associated with them. Marine unicellular cyanobacteria assigned to the genus *Synechococcus* and possessing phycoerythrin as their primary accessory light-harvesting pigment were classified as marine cluster A (MC-A) and distinguished from marine cluster B (MC-B), members of which have phycocyanin as their major light-harvesting pigment (120). MC-A *Synechococcus* and *Prochlorococcus* strains are very closely related, despite considerable differences in their light-harvesting apparatus, and represent a monophyletic clade (81, 116).

A consideration of the cyanobacterial cell envelope is important since this is the site at which the phage attaches to the cell and also the potential barrier to the injection of phage DNA. The structure of the cyanobacterial cell envelope has been reviewed by Gantt (36). The peptidoglycan layer surrounding the cytoplasmic membrane, although structurally similar to that of Gram-negative bacteria, is thicker and has a chemical composition which is closer to that of Gram-positive bacteria. The cyanobacterial outer membrane in many species is surrounded by a carbohydrate-enriched fibrous glycocalyx sheath. Some bacteria possess an S-layer outside the outer membrane. S-layers are composed of two-dimensional, monomolecular quasi-crystalline arrays of identical units of protein or glycoprotein and have been found in 60 unicellular strains and five filamentous cyanobacteria strains (104). Recently the presence of an S-layer was reported for a member of the abundant marine MC-A *Synechococcus* group (93).

Characterization and Classification of Cyanophages

The discovery and characterization of cyanophages has proceeded in two very distinct stages. Firstly, interest was focused on cyanophages from freshwater systems and there was then a period of comparatively little activity until the early 1990s when attention was shifted to marine systems. All the cyanophages so far characterized fall into the three families of tailed phages with double-stranded DNA genomes recognized by the International Committee on the Taxonomy of Viruses (ICTV): the myoviruses, siphoviruses, and podoviruses (see chapter 2 for a review of phage classificiation by virion morphology). Safferman et al. (89) similarly proposed three genera for the cyanophages, but these taxonomic suggestions have never been adopted by the ICTV. However, the non-taxonomic terms derived from the three proposed genera—cyanomyovirus, cyanopodovirus, and cyanosiphovirus (formerly cyanostylovirus)—are frequently employed as useful shorthand for discussing cyanophages.

There is a growing feeling that there are inherent problems in the phenetic approach to phage classification (58) and, indeed, a taxonomy based on sequenced genomes may thus be more appropriate (88). Certainly there is strong evidence that genetic mosaicism and access to a common phage gene pool is an important feature of cyanophage evolution (45) and, as a consequence, a phenetic classification of cyanophages is probably of very limited phylogenetic value. Consequently, while the cyanophages will be discussed here in terms of the ICTV-approved families, this should not be interpreted as implying any phylogenetic relatedness between the cyanophages classified in this way unless there is additional, nonmorphological evidence to indicate shared homologous features. Similarly, there is no attempt to assign cyanophages to the generic subdivisions of these families. What is intended instead is a discussion, here, of the diversity of the cyanophages assigned to these families. More comprehensive listings and descriptions of cyanophages can be found elsewhere (69, 110)

The myoviruses, of which the archetype is the coliphage T4 (reviewed in chapter 18), have long contractile tails and a head that exhibits icosahedral symmetry, or symmetry which is derived from a basic icosahedral structure. There is considerable morphological and molecular diversity amongst the cyanophages assigned to this family and the morphological diversity of marine cyanomyoviruses has been particularly remarked on (112, 121). One of the marine cyanomyoviruses has been reported to have neck filaments (112), which are rare amongst phages in general and had previously been reported for cyanophages infecting filamentous freshwater cyanobacteria (1, 78). Members of the family have been isolated which infect filamentous or unicellular hosts. The mol%GC content of cyanomyoviruses spans a range of 37% for phage N1 (2) to 55% for phage AS-1M (98); there is also enormous variation in genome sizes, with N1 at 37 kb (2) and the marine cyanophage S-PM2 at 196 kb (45). The large majority of phage isolates infecting MC-A *Synechococcus* strains (63, 112, 121, 128) are myoviruses, as are phages infecting *Prochlorococcus marinus* strains belonging to the low-light clade (109).

The siphoviruses are distinguished by their long, noncontractile tails and are the least frequently isolated cyanophages from both freshwater and sea water and are the least characterized. The bulk of the cyanosiphoviruses so far isolated infect unicellular hosts, although viruses assigned to this family have been reported to infect the marine filamentous cyanobacterium *Lyngbya majuscula* (47). Again, as with the cyanomyoviruses, there is considerable variation in mol%GC content, ranging from 46% for S4-L (52) to 70–74% for S1 (1). S-1 has a genome of 38 kb (1).

The podoviruses possess heads with icosahedral symmetry and short tails. The first cyanophage to be discovered, which caused lysis of several species of cyanobacteria from the genera *Lyngbya*, *Plectonema*, and *Phormidium*, hence the designation LPP-1, belongs to this family (90). Mol%GC contents show rather less variation in cyanopodoviruses, ranging from 53–55% for LPP-1 (64) to 66–67% for SM-1 (91). Genome sizes are typically around 42 kb for LPP-1 (64) to 48 kb for the marine isolate P60 (22). Phages infecting the high-light clade of *Prochlorococcus marinus* belong to this family (109).

Infection of Host Cells

Host Range

There have been many problems associated with establishing the host range of freshwater cyanophages infecting both filamentous and unicellular cyanobacteria. These problems arise largely, though not exclusively, from the previously mentioned problems with cyanobacterial taxonomy and are excellently discussed by Suttle (110). The situation with marine cyanophages is not so problematic, particularly with regard to phages infecting strains of unicellular cyanobacteria belonging to the MC-A *Synechococcus* and *Prochlorococcus* lineage. However, in spite of the taxonomic problems there are two consistent observations that can be made relating to cyanophage host range, namely that phages that infect filamentous strains do not infect unicellular strains and phages that infect marine strains do not infect freshwater strains.

Host range studies are carried out under laboratory conditions and there are a number of factors that make the extrapolation to natural assemblages problematic. In particular, natural assemblages are usually composed of different ecotypes that may vary in their sensitivity to infection by a particular phage for a variety of reasons: The physiological state of the cell or its position in the cell cycle may determine the ability of the phage to adsorb (see “Adsorption” below). The incoming phage DNA might be subject to restriction by endonucleases specified by coinfecting phages. There may, of course, be selection for mutation to resistance, though this is likely to incur a trade-off in fitness. Another reason for differential efficiency of infection might be immunity arising from lysogeny, and some lytic phages are also able to prevent secondary infection by other lytic or temperate phages by a process known as superinfection exclusion.

Detailed studies have been carried out on the host range of phages infecting MC-A *Synechococcus* strains.

Waterbury and Valois (121) found considerable variation in the host ranges of their *Synechococcus* phage isolates. Some phages would infect as many as 10 of the 13 strains tested, whereas others would infect only the strain used for isolation. One phage isolated on a MC-A *Synechococcus* strain would infect other MC-A strains, but also a MC-B strain (WH8101). None of the phages would infect the freshwater strain *Synechococcus* sp. PCC 6307. Suttle and Chan (112) isolated phages on both MC-A and MC-B hosts. Host range was not correlated with the geographical locations where the phages and hosts were isolated. Phages from all three phage families were isolated that infected MC-B strains, but all the phages capable of infecting MC-A strains were myoviruses. One of their isolates, a myovirus (S-PWM3), infected a green *Synechococcus* (presumably MC-B) as well as three MC-A strains. Again no infectivity against freshwater strains was observed. Thus, myoviruses appear to exhibit broader host ranges.

Lu et al. (63) also isolated phages against phycoerythrin-containing MC-A and phycoerythrin-lacking (presumably MC-B) hosts. Again they found that phages infecting MC-A *Synechococcus* strains had a broader host range and that there were phages infecting hosts from both MC-A and B strains. These observations have been extended by the discovery by Sullivan et al. that there are marine phages that can infect both *Prochlorococcus marinus* and *Synechococcus* (109). Furthermore, there was a clear distinction between phages which infected the strains of the high-light clade of *Prochlorococcus marinus*, which were highly strain-specific podoviruses, and those which infected strains of *Synechococcus* and the low-light clade of *P. marinus*, which were broader host range myoviruses. One caveat applying to all such studies is that they may be very sensitive to which host strain a phage was propagated on prior to assessment of host range, since the hosts may possess different restriction-modification systems.

Dutta et al. (33) have revisited the concept of the “nascent phage” based on experiments with the T4-like phage LZ4. Drawing on earlier work—including the observation that intracellular phages (T2 and T4) are associated with the cytoplasmic membrane (100) and that newly released phage carry a membrane fragment, loss of which is correlated with cofactorless infection (19)—it is proposed that nascent phages are newly synthesized phages which can infect related bacteria that lack the normal phage receptor. In this context, it was reported for cyanophage LPP-1 that newly formed phage particles were found to be closely associated with the thylakoids and to remain attached after lysis (105). If this membrane association can similarly lead to a nonspecific broadening of the host range such as that proposed for phage LZ4, then such nascent-phage status could seriously complicate assessment of cyanophage host range and ecological significance (see”Contact Rates” below).

Adsorption

The first key step in the interaction between a phage and a potential host cell is the adsorption of the phage to the cell envelope, a process which involves the phage adhesins recognizing and binding to specific cell-surface receptors, commonly either lipopolysaccharide or protein. Consequently, it is surprising that there has been so little work done on identifying the receptors recognized by cyanophage adhesins, or on characterizing the adhesins themselves. Discussion of what has been accomplished in the area of cyanophage adsorption follows.

Adsorption of phage AS-1 was found to be significantly dependent on light, though this dependence could be reduced by increasing the concentration of Na^+ ions (28). LPS has been implicated in the absorption of the cyanomyovirus AS-1 by virtue of the ability of purified polysaccharide to inactivate AS-1. AS-1 adsorption protein(s) also have been preliminarily identified by the purification of a fraction with receptor activity (92). Disruption in *Anabaena* sp. strain PCC 7120 of the genes thought to encode undecaprenyl-phosphate galactosephosphotransferase (*rfbP*) and first mannosyl transferase (*rfbZ*) led to resistance to the obligately lytic phage A-1(L), as well as the temperate phage A-4(L) (129). Both these enzymes are involved in the synthesis of the O-antigen component of LPS and electrophoretic analysis showed that the interruption of the *rfbP* and *rfbZ* genes led to a change in or loss of the characteristic pattern-length of the LPS. The O antigen is comprised of serially repeated, strain-specific oligosaccharide units. Thus, the currently available evidence implicates LPS in cyanophage adsorption and variation in the nature of the O-antigen component as a determinant in phage host range.

The nature of the cell surface is likely to be influenced by the nutrient stresses imposed on the cell, particularly with respect to proteins involved in nutrient transport. It was shown as early as 1940 for a coliphage that the adsorption rate constant under optimal growth conditions was more than 60 times greater than under poor conditions (30). Consequently both the presence and density of some receptors might be influenced by nutrient availability, as is the case for phage λ where the receptor is induced by maltose. However, the only study on the effect of nutrient starvation on cyanophage adsorption showed that phosphate depletion had no effect on adsorption kinetics (126). In contrast it has been reported in the case of a coastal strain of *Synechococcus* that phage would only bind to about 10% of the cells, suggesting that only a small proportion of the population was expressing the particular phage receptor at any particular time (110). In contrast, phage were found to attach to all cells in a clonal *Synechococcus* population (46). It has even been suggested that there may be ecological advantages in an oligotrophic environment in expressing decoy phage receptors leading to phage adsorption and DNA entry, but resistance to subsequent viral replication and lysis, thus permitting the incoming DNA to be used as a valuable source of carbon, nitrogen, and phosphorus (35).

Restriction-Modification Systems and DNA Modifications

Restriction endonucleases are thought to represent a mechanism by which bacterial cells can degrade incoming "foreign" DNA and thus resist phage infection even if the cell possesses the phage receptor. Certainly a large number of restriction-modification systems have been characterized in cyanobacteria and such an adsorption-then-restriction process would explain the difference in efficiency of plating of phage N-1 on *Anabaena* sp. PCC 7120 and *Anabaena variabilis* (29).

Restriction endonucleases encoded by phages can also represent mechanisms by which the phage can target the degradation of the genome of the infected cell and prevent superinfection by other phages lacking the appropriate modification. Phage AS-1 has been shown to encode a restriction-modification system that can degrade host DNA (113). Wilson et al. (128) also observed that *Eco*R1 digested the DNA of four phages propagated on the MC-A strain *Synechococcus* sp. WH7803, but failed to digest the DNA of a fifth, the myovirus S-WHM1, suggesting that it encoded its own restriction-modification system.

Another mechanism by which a phage can protect itself against both its own or the host's restriction enzymes is to substitute, by a modification of the biosynthetic pathway, an unusual base (e.g., hydroxymethyl cytosine) throughout the genome in place of one of the normal bases. This latter approach is adopted by phage T4 and provides almost unlimited protection against nucleases that recognize unmodified sequences. Protection against nucleases recognizing modified sequences is conferred by glucosylation of the hydroxymethyl cytosine residues (21). AS-1 DNA was shown to contain about 5% of a modified nucleotide which was not 5-methyldeoxycytidylic acid (113) and lack of modification of the replicated phage DNA early in the infection cycle may explain the phenomenon of phage-induced light (PIL) DNA (48). In addition, 2-aminoadenine has been found to substitute for adenine in the DNA of cyanophage S-2L (53, 54). Many restriction endonucleases do not digest MC-A *Synechococcus* phage DNA effectively, suggesting the presence of modified bases or the lack or restriction sites (63, 128).

There may also be a strong selection pressure on phages to lose potential restriction sites. Such may be the case for cyanophages that infect species of *Anabaena* and *Nostoc*. An analysis of restriction endonuclease cleavage of DNA isolated from these phages has provided evidence for counter-selection of restriction endonuclease sites. These include sites containing subsequences that are methylated by host (*Anabaena* sp. PCC 7120) methylase(s) (10).

Lysogeny

Although many cyanophages are obligately lytic, a significant proportion are capable of lysogeny and this topic is well established for freshwater cyanophages and was extensively reviewed by Sherman and Brown (97) (see chapter 7 for a discussion of general aspects of lysogeny). It was hoped that temperate phages would be useful as tools for fine structure genetic mapping, but this hope was never realized. An important phenomenon associated with lysogeny is that prophage may significantly alter the phenotype of the host cell, a phenomenon known as phage conversion. As yet, however, there is no clear evidence yet for cyanophage "conversion" of host strains. The earliest reports on lysogeny focused on phages infecting filamentous strains and eventually *Plectonema boryanum* and phage LPP-2SPI became an accepted system to study (79). The SPI strain of *P. boryanum* can continually liberate phages, but induction of a prophage could not be achieved with ultraviolet (UV) light, X-rays or mitomycin C, and the cells though sensitive to infection by LPP-1 were immune to LPP-2. However, Rimon and Oppenheim (83) were able to isolate a temperature-sensitive mutant of the phage that could be induced at 40 °C in the light. Inhibitor studies (25) showed that transcription and translation were required for induction, as was light.

Attempts to induce prophages in MC-A *Synechococcus* strains by a range of treatments including temperature shock, light shifts, UV and X-irradiation, and mitomycin C were not successful (121). No intact prophages have been found in the sequenced cyanobacterial genomes. However, the marine MC-A strain *Synechococcus* sp. WH8102 contains 16 putative phage integrases and three possible integrase regulators (80). *Prochlorococcus* strains MIT9313 and MED4 contain four and one, respectively (86). These phage integrase genes may represent the remnants of once fully functional prophages. Sode et al. (107) isolated a temperate phage infecting a marine *Synechococcus* strain, though not of the MC-A group, which was inducible by heavy metals, particularly copper (106).

More recently attempts have been made to detect lysogeny in natural assemblages of MC-A *Synechococcus*. One study was carried out on a *Synechococcus* bloom in a pristine fjord in British Columbia, Canada (77). The samples were incubated with and without mitomycin C and the abundances of *Synechococcus* and infectious cyanophage (using strain WH7803 as a host) were estimated. Lysogenic phage production was estimated from the difference in cyanophage abundance between the mitomycin-C-treated samples and the controls. On this basis induction of prophages occurred in 0.6% of the *Synechococcus* population. This represents a minimum value as only phage that infect strain WH7803 would be counted and mitomycin C may not be an effective inducer of prophages in *Synechococcus* spp. In a second study, McDaniel et al. (70) analyzed lysogeny using mitomycin C during an annual cycle in Tampa Bay, Florida. The frequency of lysogens was inversely correlated with *Synechococcus* abundance. Lysogens were primarily detected during the late winter months, though lysogens were also detected in August, preceding a secondary autumnal *Synechococcus* bloom.

Autoplaquing refers to the appearance of plaques within a bacterial lawn that has not been infected with phage, and is thought to be due to the spontaneous induction of a prophage. The phenomenon of autoplaque formation has been observed for about 50% of clonal *Nodularia spumigena* isolates from the Baltic Sea with either cyanomyoviruses or cyanosiphoviruses being present within the cell lysates; autoplaque formation was associated with senescent *Nodularia* cultures and cultures exposed to high light/temperature (C. Jenkins and P. K. Hayes, personal communication). Studies with the filamentous marine cyanobacterium *Phormidium persicinum* (Provasoli strain) have shown it to be a lysogen and that the prophage could be induced with mitomycin C or UV (76). This led to the suggestion that the rapid disappearance of *Trichodesmium* blooms and the lysis of laboratory cultures following exposure to stresses such as a sudden temperature increase (75) could be explained in the same way.

Pseudolysogeny

There is evidence that obligately lytic *Synechococcus* phages can enter the pseudolysogenic state (126), a phage–host relationship in which a phage-infected cell grows and divides even though its virus, though metabolically inactive, is pursuing a lytic infection. A commonly cited example is phage T3 infection of F^+ *Escherichia coli* under starvation conditions (13). When the obligately lytic phage S-PM2 (a myovirus) was used to infect *Synechococcus* sp. WH7803 cells grown in phosphate-replete or phosphate-deplete media it was found that there was an apparent 80% reduction in the burst size under phosphate-deplete conditions. However, a more detailed analysis showed that 100% of the phosphate-replete cells lysed, compared with only 9% of the phosphate-deplete cells, suggesting that the majority of phosphate-deplete cells were pseudolysogens. Similar observations were made with two other obligately lytic *Synechococcus* myoviruses, S-WHM1 and S-BM1.

Temperate phages can alter the phenotype of their lysogenic host via the phenomenon of phage conversion (see "Lysogeny" above). In the pseudolysogenic state both lytic and temperate phages might be able modify the phenotype of the host cell. There is one example in the literature (55). T7 is capable of forming a pseudolysogenic relationship with strains of *Shigella dysenteriae*. *S. dysenteriae* is normally not capable of utilizing mannitol (Mann^-) or lactose (Lac^-), but T7 pseudolysogens become Mann^+ Lac^+, which is attributed to the effect of the phage endolysin altering the permeability of the cell envelope. It may be appropriate to coin the term "pseudolysogenic conversion" to refer to the temporary phenotypic alteration in the phenotype of a pseudolysogenic host. In this context it is worth noting that the marine cyanophage P60 carries the *phoH* gene, which in other systems is associated with the host's reponse to phosphate starvation (22).

Effects on Host Cell Physiology

Effects on Photosynthesis and Respiration

A key feature of the cyanobacteria is their oxygenic photosynthesis and consequently much interest has been focused on the impact of phage infection on the host's photosynthetic metabolism and, conversely, the dependence of phage replication on photosynthesis. Obviously, photosynthesis is the primary source of energy for phage replication and assembly and for the fixed carbon for the biosynthesis of nucleotides. However, there are more subtle points at which the two processes are likely to interact. For example, phycobiliproteins represent a very substantial component of total cell protein, thus a continued synthesis of phycobiliproteins will represent a drain on resources for phage replication and existing phycobiliproteins represent a potential source of amino acids for the biosynthesis of phage proteins. However, phycobiliproteins are also components of the light-harvesting antenna for photosynthesis.

Cyanobacteria in the natural environment also are subject to a light–dark cycle and it is important to know how phage replication will be affected by the switch from light to dark metabolism (see below). Dark metabolism in cyanobacteria relies on energy production by the oxidative pentose phosphate pathway in which glucose-6-phosphate dehydrogenase is a key enzyme. There are several reports of the activity and properties of glucose-6-phosphate dehydrogenase being altered following phage infection (6, 9, 102) and its normal susceptibility to redox control has been reported to become uncoupled during phage infection of *Synechococcus* sp. PCC 6301 (27).

The photosynthetic apparatus of cyanobacteria, like the chloroplasts of higher plants, employs two photosystems: PSI and PSII (11, 41). Light energy harvested by PSII is used to drive the photolysis of water and an excited electron is transferred from chlorophyll P_{680} to pheophytin, which reduces a bound plastoquinone molecule Q_A. This in turn reduces a second quinone Q_B. From here electrons flow through the intersystem chain to PSI, concomitantly generating ATP. In PSI a second photon-induced photochemical charge separation occurs at chlorophyll P_{700} and the electron is passed through a series of electron carriers to ferredoxin and thence to $NADP^+$. According to the energy needs of the cell, electrons can also flow from ferredoxin back into the intersystem electron transport pathway to enhance ATP production, a process referred to as cyclic photophosphorylation.

An important feature of the photosynthetic process that would affect phage replication is the fact that the water-splitting photochemistry of PSII produces various radicals and toxic oxygen species that cause damage to PSII. At the core of PSII lies a heterodimer of two related proteins, D1 and D2, which binds the pigments and cofactors necessary for primary photochemistry. During active photosynthesis D1, and to a lesser extent D2, turns over rapidly as a result of photodamage and is replaced by newly synthesized polypeptides in a repair cycle. When the rate of photoinactivation and damage of D1 exceeds the capacity for repair, photoinhibition occurs, resulting in a decrease in the maximum efficiency of PSII photochemistry (8).

There are two aspects of photoinhibition that are relevant to a consideration of cyanophages. Firstly, the production of radicals and toxic oxygen species may directly damage components of the nascent phage. Secondly, a functional PSII may be required to provide the energy for the process of phage replication. It should be remembered that cyanobacteria in the natural environment are likely to be nutrient stressed and this will significantly enhance the likelihood of photoinhibition. Phage T4 possesses a variety of mechanisms that ensure host gene expression is shut down shortly after infection. Were cyanophages to adopt a similar strategy, then the PSII repair cycle would cease with the failure to synthesize new undamaged D1. The central importance of this particular topic is made clear by the observation that genomic studies on a marine cyanomyovirus S-PM2 indicate that it encodes the D1 and D2 proteins of PSII (67), which suggests that the phage may ensure a continued repair cycle for PSII (see below). The occurrence of genes encoding D1 and D2 proteins is widespread amongst independent marine cyanomyoviruses isolated from geographically distinct provinces (A. Millard, personal communication). This is not a universal strategy, however, since the marine cyanopodovirus P60, which has a much smaller genome that has been fully sequenced, does not encode D1 or D2.

Many of the early studies on the relationship between phage infection and photosynthesis relied heavily on the use of inhibitors such as DCMU and CCCP, whose mode of action was not clearly established at the time. CCCP is a protonophoric uncoupler, which blocks both photosynthetic and respiratory ATP production, but is also now known to inhibit intersystem photosynthetic electron transport by causing futile cycling of electrons around PSII (94). DCMU inhibits photosynthetic electron transport at the acceptor side of PSII by binding to the D1 protein of PSII competing with quinone for binding to the Q_B site and thereby inhibiting electron transport from the primary quinine acceptor, Q_A, to the secondary quinone acceptor, Q_B. However, this inhibition of photosynthetic electron transport will have a number of pleiotropic effects which make it difficult to establish whether any effects of DCMU are due to its primary effect on electron transport or its secondary effects on other processes such as redox signaling.

In this context it is important to mention that DCMU can effect the transcription of the multiple *psbA* genes (encoding D1 isoforms) in cyanobacteria (95, 103) and may have a more general effect on the translation machinery (95). DCMU, somewhat counterintuitively, inhibits D1 turnover in cyanobacteria (20, 42), but nevertheless has a protective effect on PSII, reducing photoinhibition, probably by altering the conformation of the Q_B site of D1 such that acceptor-side photoinhibition does not occur (74). It is in the light of these concerns about D1 turnover and the primary and secondary effects of DCMU that early experiments on the relationship between phage infection and photosynthesis should be considered.

In the case of phages infecting unicellular cyanobacteria, the general trend is for there to be a variable but usually substantial dependence of the infection cycle on the continued photosynthetic activity of the infected host. An almost complete dependence of phage replication on photosynthesis was observed with phage AS-1 (a lytic myovirus), which infects the unicellular cyanobacterium *Synechococcus* sp. strain PCC 6301 (formerly *Anacystis nidulans*) (4). Darkness, post-infection, did not completely abolish phage replication, though the yield was reduced to 2% of that obtained with light-incubated cells. DCMU reduced the phage yield by 73%, for which the simplest explanation is that cyclic photophosphorylation via PSI could support a reduced level of phage production. The uncoupler CCCP, not surprisingly, completely abolished phage production.

Cyanophage AS-1 was found to progressively inhibit oxygen evolution in infected cells, with a rate of less than 10% that of uninfected cells at the time that lysis was commencing. This was coupled with a progressive reduction in PSII-mediated photosynthetic electron flow, whereas the activity of PSI was unaffected by the infection (114). The target of PSII inhibition is probably at the level of the secondary acceptor, Q_B (114). However, perhaps the most significant observation was that the turnover of the D1 protein was inhibited in infected cells (114). This suggests that the progressive impairment of photosynthesis as reflected by oxygen evolution was due to accumulating photodamage and consequent photoinhibition of PSII. Thus, the infected cells presumably lacked the normal D1 repair cycle of uninfected cells and AS-1 was not providing a phage-encoded repair cycle.

A very different phage–host relationship was observed following infection of *Microcystis aeruginosa* strain NRC-1 with the cyanopodovirus SM-1 (65). Photoassimilation of carbon dioxide was not significantly inhibited prior to lysis of the cells, suggesting that the cells were not sensitive to photoinhibition and must have a functioning PSII repair cycle. Furthermore, withdrawal of carbon dioxide, absence of light or presence of DCMU (10^{-6} M) completely abolished phage replication and even led to the loss of infectious centers. DCMU, which permitted a degree of phage replication in the case of AS-1, may have exerted such a strong effect with SM-1 because the diversion of electrons into cyclic photophosphorylation would have prevented the production of NADPH required for carbon dioxide fixation and phage replication. Thus, replication of SM-1 was completely dependent on a fully functional photosynthetic apparatus and could not be even partially sustained by cyclic phosphorylation or dark respiration. It should be noted that there is considerable taxonomic confusion surrounding strain NRC-1, which is well discussed by Suttle (110); it likely corresponds to *Synechococcus* sp. PCC 6911.

The third cyanophage infecting unicellular cyanobacteria for which such studies have been carried out is AS-1M (a lytic myovirus) (96). When *Synechococcus cedrorum* (= *Synechococcus* sp. strain PCC 6908) was infected with cyanophage AS-1M, carbon dioxide fixation and oxygen evolution continued at high levels throughout the latent period, declining only immediately prior to lysis, as was the case with SM-1. The similarity to SM-1 infection extended to the complete lack of phage replication in the presence of DCMU, or the absence of carbon dioxide or light. However, the presence of exogenous glucose in the presence of DCMU, or in the dark, restored phage yields to about 10% of the light-incubated controls, which suggests a respiratory production of NADPH given light plus DCMU and a respiratory production of ATP given dark plus glucose.

Studies with filamentous cyanobacteria, as opposed to unicellular strains, have led to the observation of somewhat different relationships between phage infection and photosynthesis. Cyanophage N-1 (a myovirus) infection of *Nostoc muscorum* showed that phage replication and release was largely dependent on photosynthetic metabolism throughout infection rather than there being a particular critical time during which light was required for the phage replication cycle (3). carbon dioxide fixation began to decline some halfway through the infection cycle, however, and was absent immediately prior to lysis. Use of the inhibitor DCMU at a concentration (10^{-6} M) which completely inhibited carbon dioxide fixation led to burst sizes that were 25% of control values. This result suggests that cyclic photophosphorylation alone can sustain the replicative cycle. Cyanophage N-1 infected cells have also been shown to have a reduced ferredoxin: NADP$^+$ oxidoreductase activity, which may be associated with cyclic photophosphorylation (51). Cells with DCMU (10^{-5} M) in dark still yielded about 2% of phage production compared with uninhibited cells in the light, thus indicating that oxidative phosphorylation could still support phage replication albeit at very reduced levels. This dark production of phage was completely abolished by the uncoupler CCCP.

In keeping with these observations, photosynthetic oxygen evolution was found to begin to decline shortly after infection, which may be a reflection of an impaired turnover of the PSII D1 protein leading to photoinhibition. In contrast, respiratory oxygen consumption was found to be unaffected or even increased and there was a transient increase in the activity of the key respiratory enzyme, glucose-6-phosphate dehydrogenase (6, 102), that was accompanied by a decline in glycogen stores (102). Phycocyanin was also progressively degraded during the course of N-1 infection (6, 51). This loss of phycocyanin could be interpreted as a viral strategy to reduce photon pressure on an already inhibited PSII, thus minimizing further inhibition and also reducing the potential for production of toxic oxygen species. Another, not necessarily exclusive possibility is the use amino acids, produced by such degradation, for phage protein synthesis (6, 51). A markedly different effect on carbon dioxide fixation was observed following infection of *Plectonema boryanum* with phage GIII, which led to a rapid and complete cessation of fixation (39).

A complicated pattern emerges when comparing the same phage infecting different hosts, or different phages infecting the same host. In the case of LPP-1 infecting the host *Plectonema boryanum*, Sherman and Haselkorn (99) showed that even in the dark a burst size of 15% of that observed with light-incubated controls could be obtained. The PSII inhibitor DCMU reduced the burst size by 60–70% indicating that, as in the case of N-1, cyclic photophosphorylation alone could support the replicative cycle, albeit at a reduced level. Very different results were observed when the same phage, LPP-1, was used to infect a different host, *Phormidium uncinatum* (14). The phage yield was unaffected by treatment of infected cells with DCMU or by transfer to the dark. PSII activity began to decline, presumably due to photoinhibition, 9 hours post-infection.

In contrast to the interaction of LPP-1 with *Plectonema boryanum*, infection by phage LPP-1(G) was independent of photosynthesis (38). Cyanophage LPP-1(G) was produced at the same yield in heterotrophic conditions (dark, glucose) as in photoautotrophic conditions, though aerobiosis was required for cyanophage replication in the dark. Exogenous glucose was not required for the cyanophage replication in the dark in heterotrophically grown cells. In photoautotrophically grown cells, the maximum burst size in the dark and with glucose was delayed for a period corresponding to glucose-uptake induction. Of the photosynthesis parameters tested, only carbon dioxide photoassimilation was affected during cyanophage LPP-1(G) infection under photoautotrophic conditions.

As a postscript it is worth noting the interesting link between phage infection and thylakoid biogenesis. In *E. coli*, expression of the protein PspA (phage shock protein) is strongly induced by phage infection (34) and a homolog of PspA, VIPP1, is essential for thylakoid development in *Arabidopsis thaliana* (57). In two freshwater cyanobacteria, *Synechocystis* and *Anabaena*, both a *VIPP1* and a *pspA* gene is present, and the VIPP1 gene is essential to thylakoid development as in *A. thaliana* (123). The *VIPP1* gene is thought to have originated from a gene duplication of *pspA* and thereafter acquired its new function that involves a C-terminal extension—that discriminates VIPP1 proteins from PspA—which is important for its function in thylakoid formation (123).

Other Aspects of Metabolism

Among the other areas of host physiology that are affected by phage infection is nutrient transport. Blaska et al. detected minimal uptake of glucose and glucose-6-phospate by uninfected cells of *Anacystis nidulans*, but uptake increased markedly after AS-1 infection, peaking 6 hours post-infection (16, as cited by 69). The host cell's response to energy limitation may also be affected. *A. nidulans* normally accumulates guanosine 3′-diphosphate-5′-diphosphate (ppGpp) upon shift from light to dark or upon treatment with uncouplers. This response, however, is abolished following AS-1 infection, indicating that the cell's normal control over its response to energy starvation has been lost (17).

Aspects of nitrogen and nucleic acid metabolism may also be influenced by cyanophage infection. Obviously there is a demand for nucleotides for phage replication and amino acids for phage protein biosynthesis. Rimon and Oppenheim (84) have implicated phage genes in the shutoff of host protein synthesis during LPP-2SPI infection. The need for nucleotides can, in part, be met by the degradation of host nucleic acids and in the case of AS-1-infected *A. nidulans* there was a 15- to 20-fold increase in DNase and RNase activities at different times post-infection (115). However, in AS-1-infected cells there was found to be a 4-fold increase in DNA content compared with uninfected cells, implying that breakdown products are inadequate to meet the total biosynthetic requirements. The marine cyanophage P60 has been shown to encode several enzymes involved in nucleotide biosynthesis (22), suggesting that it enhances or modifies the infected cell's biosynthesis of nucleotides. Macromolecular processes may also be affected and AS-1 infection has been associated with an inhibition of post-maturational cleavage of 23S rRNA (18).

The nitrogen economy of the infected host may be altered and, in the case of LPP-1 infection of *Phormidium uncinatum*, a steep rise in the activity of nitrate reductase can be detected both in the light and in the dark (15). It has been reported that sequences homologous to the *Klebsiella pneumoniae nifH D* and *K* genes were detected by dot-blot hybridization in the genome of temperate phages of the NP-1T series that infect *Nostoc* sp. 39 (73). Furthermore, lysogens of *Nostoc* sp. 39 carrying these temperate prophages exhibited considerably enhanced nitrogenase activity. These surprising observations are supported by the sequencing of the genome of phage AN15, which has revealed the presence of a *nif* gene (A. Baker, W. H. Wilson and D. G. Adams, personal communication), though the physiological significance of these phage-encoded *nif* genes has yet to be established.

Ecological Significance

Contact Rates

It is axiomatic that the frequency with which phages encounter and infect susceptible hosts in the natural environment will determine whether phages exert a significant selection pressure on the hosts, in terms of both overall abundance and genetic diversity (for additional discussion on this quantitative and qualitative impact of phages on bacteria, see chapters 5 and 33). By far the most work on estimating cyanophage–host contact rates has been done with cyanophages infecting MC-A *Synechococcus* strains. The assumptions underlying the different approaches and the sometimes contrasting predictions have been discussed by Mann (66).

Waterbury and Valois (121) used a theoretical approach based on the diffusivity of phage particles and physical approaches to coagulation, coupled with observations on phage and host population densities, to estimate contact rates between phage and marine *Synechococcus* hosts in Woods Hole Harbor. On this basis it was calculated that between 0.005% (at the end of the spring bloom) and 3.2% (during a *Synechococcus* peak in July) of the *Synechococcus* population was contacted, and assumed to be infected, on a daily basis. Contacts rates have also been calculated using a model for the interaction of viruses with hosts based on diffusive transport that allows inclusion of the motion of water and host. When coupled to estimation of phage decay rates this approach has led to somewhat higher estimates of contact rates. If a burst size of 50 was assumed, as many as 33% of the *Synechococcus* population would have to have been lysed daily at one of the sampling stations (111). A subsequent study using the same approach (37) yielded figures for the proportion of the *Synechococcus* community infected ranging from 1–8% for offshore waters. In nearshore waters only 0.01–0.02% of the *Synechococcus* were lysed on a daily basis. In all cases the efficiency of infection was very low, with only 1.01–3.18% of contacts leading to infection.

The observations for nearshore waters are in good agreement with those of Waterbury and Valois (121). However, considerations of contact rates between phage and host, phage decay rates, and phage production become infinitely more complicated following the discovery that there are phages which can infect both MC-A *Synechococcus* and *Prochlorococcus marinus* (109), except in regions of the oceans where the temperature falls below the threshold for the occurrence of *P. marinus*. The complexity may be further enhanced if nascent cyanophages do have a greater host range than mature phage particles (discussed above).

Another complication in the calculation of contact rates in natural assemblages arises from the accurate estimation of the true number of infectious phages. Assays with a single host are likely to lead to an underestimation of infectious phage. Another significant factor is the impact of solar radiation on the infectivity of natural cyanophage assemblages. Decay rates as high as 0.75 per day have been measured for phages infecting MC-A *Synechococcus* in the surface mixed layer of the Gulf of Mexico (37). The UV-B component of sunlight leads to phage inactivation primarily by the formation of pyrimidine dimers and much of this damage may be repaired by photoreactivation involving a post-infection host cell repair mechanism. This is a well-established phenomenon in cyanophage–host systems (5, 7, 49, 59, 101) and such processes also occur in natural communities (122). The enhanced ability of a host to reactivate damaged phage if the host cell itself has been irradiated (Weigle reactivation) is also likely to be important in natural assemblages and has also been reported to occur in cyanobacteria (59). The use of unirradiated hosts to assess cyanophage abundance in natural environments, particularly those with high insolation rates, consequently may lead to a significant underestimate of cyanophage abundance.

In order to assess the impact of phage on host dynamics and diversity it is important to establish at what point the contact rate becomes a significant selection pressure on a population, leading to either the succession of intrinsically resistant strains or the appearance of resistant mutants. This threshold, which would presumably represent the point at which the phages begin to exert a significant selection pressure on the host, is the point at which the product of contact rate (assumed to be equal to the infection rate) and the burst size equals the product virus decay rate and the virus population density (66). Taking a range of experimentally determined values, the threshold would occur at between 10^2 and 10^4 cells ml^{-1}. This is in agreement with data from natural *Synechococcus* populations that suggest a genetically homogeneous population would start to experience significant selection pressure when it reached a density of between 10^3 (111) and 10^4 cells ml^{-1} (87, as cited by 110). The use of fluorescently labeled phage probes has also shown that phages present at comparatively low abundance can control microbial community structure (46). Such ideas cannot be easily extended to filamentous or colonial cyanobacteria where phage release could occur within the filament or colony.

Impact on Natural Assemblages

Much of the original interest in cyanophages arose from their potential as agents to control nuisance blooms of freshwater cyanobacteria. Cyanobacterial blooms can cause economic and social impacts by a variety of mechanisms including toxin production, impairment of water treatment processes, and eutrophication. Attempts to use cyanophages to prevent or limit blooms of cyanobacteria have been reviewed by Martin and Benson (69) (see also chapter 48, which reviews the related phenomenon known as phage therapy). Some control of *Plectonema boryanum* populations in outdoor pond facilities was achieved with phage LPP-1 (31, as cited by 69), though the phage was most effective when added prior to the formation of the bloom. However, the continued occurrence of nuisance blooms and the lack of use of cyanophages to control them is a testament to the lack of success with this approach.

There are, however, continuing reports of the termination of freshwater and marine blooms being associated with phages. The involvement of cyanophages was suggested to be a factor in the decline of *Aphanezomenon flos-aquae* blooms in shallow, eutrophic lakes in Manitoba (26). van Hannen (117) described the growth of a cyanobacterial bloom, dominated by *Oscillatoria limnetica* and *Prochlorothrix hollandica*, in two laboratory-scale enclosures of water from the shallow, eutrophic Lake Loosdrecht (the Netherlands), and attributed the population collapse of the filamentous *O. limnetica* to viral lysis. In a follow-up to these observations, Gons et al. (44) described how in similar laboratory-scale enclosures of water from the same lake, the predominating filamentous cyanobacteria grew vigorously for 2 weeks, but then their populations simultaneously collapsed, whereas coccoid cyanobacteria and eukaryotic algae persisted. The collapse coincided with a short peak in the counts of virus-like particles and transmission electron microscopy showed myoviruses, with isometric heads of about 90 nm outer diameter and tails more than 100 nm long, that occurred free, attached to and emerging from cyanobacterial cells. Expansive blooms of the toxic cyanobacterium *Lyngbya majuscula* were observed in two shallow-water regions of

Moreton Bay, Australia, that were subject to rapid bloom declines (8 to $<1\ \text{km}^2$ in <7 days) (47). Virus-like particles produced by decaying *L. majuscula* were observed using electron microscopy and appeared to be siphoviruses. The induction of temperate prophages has been proposed as the underlying mechanism of bloom termination of the ecologically important marine diazotrophic cyanobacterium, *Trichodesmium* sp. (75).

The most direct method of detecting cyanobacterium–phage interactions is to use transmission electron microscopy to determine the proportion of cells that contain visible mature phages. An assessment of the contribution of phage to host mortality can then be made based on estimates of the proportion of the infection cycle during which phage particles are visible and the assumption that mortality due to infection is twice the number of infected cells. This approach was used with marine *Synechococcus* by Proctor and Fuhrman (82). They found that, depending on the sampling station, between 0.8% and 2.8% of cyanobacterial cells contained mature phage and, assuming that phage particles were only visible for 10% of the infection cycle, it was estimated that the percentage of infected cells was actually 10-fold greater than the observed frequency. Thus, viral infection could account for as much as 56% of marine *Synechococcus* mortality. However, many of the underlying assumptions in this approach have been questioned (66, 110).

Much attention these days is focused on the questions of whether phage affect the genetic diversity of their hosts and whether this in turn will affect phage diversity. In the context of this idea, different strains of the same species that are sensitive to infection by different phages and that display different growth rates in a particular set of environmental conditions could be treated as different species in terms of competition and succession. Support for the idea of a correlation between host and phage genetic diversity comes from studies on phages infecting marine MC-A *Synechococcus* strains. Maximum *Synechococcus* myovirus diversity in a stratified water column was correlated with maximum *Synechococcus* population density (127) and changes in phage clonal diversity were observed from the surface water down to the deep chlorophyll maximum in the open ocean (130). Distinct cyanomyovirus population structures were found in estuarine water versus open ocean (130). Temporal changes in the relative abundance of specific cyanophage g20 genotypes was observed during the summer months in Rhode Island coastal water (68). All this evidence suggests the presence of different host ecotypes in each environment and a dynamic interaction between cyanophage and host.

In a study of an oligotrophic environment (Gulf of Aqaba, Red Sea) over an annual cycle, the *Synechococcus* diversity was monitored using the *rpoC1* gene (M. Mühling, N. Fuller, A. Millard, D. J. Scanlan, A. Post, W. H. Wilson, D. Marie, and N. H. Mann, unpublished results). There was considerable diversity in spring, with as many as 12 *Synechococcus* genotypes present, and this was followed by a marked decline in diversity toward the summer and autumn, when only one or two genotypes dominated, respectively. In the following winter months there was an even greater *Synechococcus* diversity than there was in spring, with as many as 24 genotypes present. The genetic diversity in the co-occurring cyanophage population was monitored using the cyanomyovirus g20. The seasonal changes in cyanomyovirus diversity paralleled that of *Synechococcu*s, with the greatest diversity in spring (28 genotypes) and winter (22 genotypes). However, cyanophage diversity was not reduced as much as *Synechococcus* diversity during the summer (17 genotypes) and autumn (15 genotypes). Given the short half-life of phages in the surface layers, the presence of cyanophage of multiple genotypes in the water column at times when the host population was dominated by only one or two genotypes indicates that cyanophage of more than one genotype were capable of infecting the dominant *Synechococcus* strains. However, the most abundant and second most abundant cyanophages (based on g20 clones) showed, respectively, a parallel pattern of abundance to that of the most abundant and second most abundant *Synechococcus* clones, which dominated the summer and autumn maxima, suggesting a specific cyanophage–host relationship.

Horizontal Gene Transfer

In addition to the original interest in cyanophage as agents of biological control there was the hope for their potential as genetic tools, in same way as phages such as λ and P1 have contributed to the elucidation of *E. coli* genetics. Unfortunately, as with biological control, cyanophages as genetic tools have not fulfilled early hopes. However, there is growing evidence, almost entirely from studies with marine cyanophage host systems, that cyanophages are important vectors of a variety of forms of horizontal gene transfer in natural assemblages. Horizontal gene transfer in general is a major factor in microbial evolution (56) and the phage-mediated transduction of both chromosomal and plasmid DNA has been reported in a variety of aquatic environments (71). Potentially phage can act not only as vectors to transfer DNA between hosts via generalized and specialized transduction, but potentially can recombine with other phages during mixed infections. (See chapter 33 for broader discussion of horizontal transfer between bacteria found within marine environments.) Phage might also acquire genes from their hosts, which could either alter their properties during a lytic infection or lead to phage conversion during lysogen formation (see chapters 27 and 47 for further discussion of lysogenic conversion).

Some of the clearest evidence for horizontal gene transfer has come from recent studies on the genomes of marine *Synechococcus* and *Prochlorococcus* strains (32, 80, 86). For the marine phages to contribute to such transfer they must be able either to package host DNA into the phage capsid or incorporate host genes into the phage genome. Evidence has been obtained, in the case of the cyanomyovirus S-PM2, that approximately 1 in 10^5 phage particles contain a host marker gene in their capsids (24). The genome of *Prochlorococcus* SS120 lacks the gene for deoxyribopyrimidine photolyase, which is present in other cyanobacteria (32). Instead, it encodes a pyrimidine dimer-specific glycosylase, which is absent in other cyanobacteria and is speculated to have been acquired from a phage genome. A similar reasoning applies to the gene encoding a type I restriction-modification system.

Although the genome of *Synechococcus* sp. WH8102 does not currently contain any prophages, there is evidence that is has been extensively altered in the past via horizontal gene transfer, probably involving phages (80). The genome contains 16 probable or possible phage integrase genes and three putative integrase regulators that may represent the fossil remnants of prophages in the ancestral genomes. There are also putative integrases in genomes of the *Prochlorococcus* strains MIT9313 and MED4 (86). Many of the multiple putative phage integrases in *Synechococcus* sp. WH8102 occur in regions of the genome with an anomalously low mol%GC content and an atypical trinucleotide composition suggestive of genomic islands. Several of the genes found in these potential genomic islands appear to be involved in the carbohydrate modification of the cell envelope, for example glycosyltransferases and enzymes involved in the synthesis of sialic acid (80). The nature of the cell surface is a critical determinant in phage attachment and susceptibility to grazing and thus is a target for intense selection pressure. One possibility is that *Synechococcus* sp. WH8102 may use these enzymes to modify the cell envelope so as to evade grazers or phages. Support for this idea comes from the observations that several of the regions of the genomes of marine *Synechococcus* and *Prochlorococcus* strains, which are suggested to have been acquired by horizontal gene transfer or lost by deletion events, encode genes involved in lipopolysaccharide and/or surface polysaccharide biosynthesis (86). The discovery of cyanomyoviruses that can infect strains of *Synechococcus* and *P. marinus* extends the potential for horizontal gene transfer between these genera.

Horizontal gene transfer may also involve recombination between phages or the acquisition of host genes by phages. There is clear evidence from the sequencing of the marine cyanomyovirus S-PM2 that their genomes encode a number of genes whose closest homologs are in cyanobacterial genomes (this laboratory, unpublished results). Furthermore, S-PM2 encodes the *psbA* and *psbD* genes of PSII, which must have been acquired from a host *Synechococcus* (67).

Conclusions

All the cyanophages so far isolated fall into just three of the recognized families of phages: the tailed phages with double stranded DNA genomes. However, it is impossible to generalize about cyanophages as they exhibit enormous diversity in terms of morphology, genome size, mol% GC composition, genetic relatedness and the effects of infection on host cell physiology. There are several important questions remaining to be answered about cyanophages and the impact on their cyanobacterial hosts. Genomic studies on marine cyanophages are beginning to reveal fascinating insights into both infection strategies and cyanophage evolution. The publication of the genomes of some freshwater cyanophages will extend our knowledge in these areas, and, importantly, will clarify the evolutionary relationships between freshwater and marine cyanophages.

An accurate assessment of their ecological significance will depend on a number of factors. There is still no methodology to estimate the true abundance of infectious cyanophage in a particular environment. Our understanding of the selection pressure of cyanophages on genetic diversity and succession in natural cyanobacterial assemblages, together with their contribution to biogeochemical cycles, is very limited. Topics such as the significance of lysogeny, pseudolysogeny and horizontal gene transfer have scarcely begun to be investigated. In short, the study of cyanophages has a long way to go.

References

1. Adolph, K. W., and R. Haselkorn. 1973. Isolation and characterization of a virus infecting a blue-green alga of the genus *Synechococcus*. Virology 54: 230–236.
2. Adolph, K. W., and R. Haselkorn. 1971. Isolation and characterization of a virus infecting the blue-green alga *Nostoc muscorum*. Virology 46:200–208.
3. Adolph, K. W., and R. Haselkorn. 1972. Photosynthesis and the development of the blue-green algal virus N-1. Virology 47:370–374.
4. Allen, M. M., and F. Hutchison. 1976. Effect of some environmental factors on cyanophage AS-1 development in *Anacystis nidulans*. Arch. Microbiol. 110:55–60.
5. Amla, D. V. 1979. Photoreactivation of ultraviolet irradiated blue-green alga: *Anacystis nidulans* and cyanophage AS-1. Arch. Virol. 59:173–179.
6. Amla, D. V., P. Rowell, and W. D. P. Stewart. 1987. Metabolic changes associated with cyanophage N-1 infection of the cyanobacterium *Nostoc muscorum*. Arch. Microbiol. 148:321–327.

7. Asato, Y. 1976. Ultraviolet light inactivation and photoreactivation of AS-1 cyanophage in *Anacystis nidulans*. J. Bacteriol. 126:550–552.
8. Baena-Gonzalez, E., and E. M. Aro. 2002. Biogenesis, assembly and turnover of photosystem II units. Phil. Trans. R. Soc. Lond. B Biol. Sci. 357:1451–1459.
9. Balogh, A., G. Borbely, C. Cseke, J. Udvardy, and G. L. Farkas. 1979. Virus infection affects the molecular properties and activity of glucose-6-P dehydrogenase in *Anacystis nidulans*, a cyanobacterium. Novel aspect of metabolic control in a phage-infected cell. FEBS Lett. 105:158–162.
10. Bancroft, I., and R. J. Smith. 1988. An analysis of restriction endonuclease sites in cyanophages infecting the heterocystous cyanobacteria *Anabaena* and *Nostoc*. J. Gen. Virol. 69:739–743.
11. Barry, B. A., R. J. Boerner, and J. C. de Paula. 1994. The use of cyanobacteria in the study of the structure and function of photosystem II, pp. 217–257. *In* D. A. Bryant (ed.) The Molecular Biology of Cyanobacteria. Kluwer Academic, Dordrecht.
12. Bergh, O., K. Y. Borsheim, G. Bratbak, and M. Heldal. 1989. High abundance of viruses found in aquatic environments. Nature 340:467–468.
13. Birge, E. A. 2000. Bacterial and Bacteriophage Genetics. Springer, New York.
14. Bisen, P. S., S. Audholia, H. D. Shukla, A. Gupta, and D. P. Singh. 1988. Evidence for photosynthetic independence of viral multiplication in cyanophage LPP-1 infected cyanobacterium *Phormidium uncinatum*. FEMS Microbiol. Lett. 52:225–228.
15. Bisen, P. S., S. N. Bagchi, and S. Audholia. 1986. Nitrate reductase activity of a cyanobacterium *Phormidium uncinatum* after cyanophage LPP-1 infection. FEMS Microbiol. Lett. 33:69–72.
16. Blashka, K. H., L. Hwang-Lee, G. Cohn, J. Blamire, and R. E. McGowan. 1982. Altered metabolite incorporation into the cyanobacterium *Anacystis nidulans* as a result of cyanophage AS-1 infection. Microbios 34:141–152.
17. Borbely, G., C. Kaki, A. Gulyas, and G. L. Farkas. 1980. Bacteriophage infection interferes with guanosine 3′-diphosphate-5′-diphosphate accumulation induced by energy and nitrogen starvation in the cyanobacterium *Anacystis nidulans*. J. Bacteriol. 144:859–864.
18. Borbely, G., M. Kolcsei, and G. L. Farkas. 1976. The postmaturational cleavage of 23 S ribosomal RNA in *Anacystis nidulans* is inhibited by infection with cyanophage AS-1. Mol. Biol. Rep. 3:139–142.
19. Brown, D. T., and T. F. Anderson. 1969. Effect of host cell wall material on the adsorbability of cofactor-requiring T4. J. Virol. 4:94–98.
20. Brusslan, J., and R. Haselkorn. 1989. Resistance to the photosystem-II herbicide diuron is dominant to sensitivity in the cyanobacterium *Synechococcus* sp. PCC7942. EMBO J. 8:1237–1245.
21. Carlson, K., E. A. Raleigh, and S. Hattman. 1994. Restriction and modification, pp. 369–381. *In* J. D. Karam (ed.) Bacteriophage T4. ASM Press, Washington, D.C.
22. Chen, F., and J. Lu. 2002. Genomic sequence and evolution of marine cyanophage P60: a new insight on lytic and lysogenic phages. Appl. Environ. Microbiol. 68:2589–2594.
23. Chisholm, S. W., S. L. Frankel, R. Goericke, R. J. Olson, B. Palenik, J. B. Waterbury, L. Westjohnsrud, and E. R. Zettler. 1992. *Prochlorococcus marinus* nov. gen. nov. sp.: an oxyphototrophic marine prokaryote containing divinyl chlorophyll *a* and chlorophyll *b*. Arch. Microbiol. 157:297–300.
24. Clokie, M., and N. H. Mann. 2003. Encapsidation of host DNA by bacteriophages infecting marine *Synechococcus* strains. FEMS Microb. Ecol. 46:349–352.
25. Cocito, C., and D. Goldstein. 1977. Inhibition of lytic induction in lysogenic cyanophyces. J. Virol. 23:483–491.
26. Coulombe, A. M., and G. G. C. Robinson. 1981. Collapsing *Aphanizomenon flos-aquae* blooms: possible contributions of photo-oxidation, O_2 toxicity, and cyanophages. Can. J. Bot. 59:1277–1284.
27. Cseke, C., A. Balogh, and G. L. Farkas. 1981. Redox modulation of glucose-6-P dehydrogenase in *Anacystis nidulans* and its uncoupling by phage infection. FEBS Lett. 126:85–88.
28. Cseke, C. S., and G. L. Farkas. 1979. Effect of light on the attachment of cyanophage AS-1 to *Anacystis nidulans*. J. Bacteriol. 137:667–669.
29. Currier, T. C., and C. P. Wolk. 1979. Characteristics of *Anabaena variabilis* influencing plaque formation by cyanophage N-1. J. Bacteriol. 139:88–92.
30. Delbrück, M. 1940. Adsorption of bacteriophage under various physiological conditions of the host. J. Gen. Physiol. 23:631–642.
31. Desjardins, P. R., and G. B. Olson. 1983. Viral control of nuisance cyanobacteria (blue-green algae), pp. 1–35. California Water Resources Center Contribution No. 185.
32. Dufresne, A., M. Salanoubat, F. Partensky, F. Artiguenave, I. M. Axmann, V. Barbe, S. Duprat, M. Y. Galperin, E. V. Koonin, F. Le Gall, K. S. Makarova, M. Ostrowski, S. Oztas, C. Robert, I. B. Rogozin, D. J. Scanlan, N. T. de Marsac, J. Weissenbach, P. Wincker, Y. I. Wolf, and W. R. Hess. 2003. Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxyphototrophic genome. Proc. Natl. Acad. Sci. USA 100:10020–100025.
33. Dutta, G., N. Hoyle, M. Robison, M. Dyen, P. Raya, B. Guttman, A. Brabban, and E. Kutter. 2003. Do "nascent" phage explain observed anomalies of T4-like LZ4? Abstracts of the 15th Evergreen International Phage Biology Meeting, Olympia, Wash., p. 21.
34. Elderkin, S., S. Jones, J. Schumacher, D. Studholme, and M. Buck. 2002. Mechanism of action of the *Escherichia coli* phage shock protein PspA in repression of the AAA family transcription factor PspF. J. Mol. Biol. 320:23–37.
35. Fuhrman, J. A. 1999. Marine viruses and their biogeochemical and ecological effects. Nature 399:541–548.
36. Gantt, E. 1994. Supramolecular membrane organizarion, pp. 119–138. *In* D. A. Bryant (ed.) The Molecular Biology of Cyanobacteria. Kluwer Academic, Dordrecht.
37. Garza, D. R., and C. A. Suttle. 1998. The effect of cyanophages on the mortality of *Synechococcus* spp. and selection for UV resistant viral communities. Microb. Ecol. 36:281–292.

38. Ginzberg, D., E. Padan, and M. Shilo. 1976. Metabolic aspects of LPP cyanophage replication in the cyanobacterium *Plectonema boryanum*. Biochim. Biophys. Acta 423:440–449.
39. Ginzburg, D., E. Padan, and M. Shilo. 1968. Effect of cyanophage infection on CO_2 photoassimilation in *Plectonema boryanum*. J. Virol. 2:695–701.
40. Goericke, R., and N. A. Welschmeyer. 1993. The marine prochlorophyte *Prochlorococcus* contributes significantly to phytoplankton biomass and primary production in the Sargasso Sea. Deep-Sea Res. A 40:2283–2294.
41. Golbeck, J. H. 1984. Photosystem I in cyanobacteria, pp. 319–360. *In* D. A. Bryant (ed.) Molecular Biology of Cyanobacteria. Kluwer Academic, Dordrecht.
42. Goloubinoff, P., J. Brusslan, S. S. Golden, R. Haselkorn, and M. Edelman. 1988. Characterization of the photosystem-II 32-kDa protein in *Synechococcus* PCC7942. Plant Mol. Biol. 11:441–447.
43. Golubic, S., T. Le Campion-Alsumard, and S. E. Campbell. 1999. Diversity of marine cyanobacteria. Bull. Inst. Oceanogr. Monaco 19:53–76.
44. Gons, H. J., J. Ebert, H. L. Hoogveld, L. van den Hove, R. Pel, W. Takkenberg, and C. J. Woldringh. 2002. Observations on cyanobacterial population collapse in eutrophic lake water. *Antonie van Leeuwenhoek* 81:319–326.
45. Hambly, E., F. Tetart, C. Desplats, W. H. Wilson, H. M. Krisch, and N. H. Mann. 2001. A conserved genetic module that encodes the major virion components in both the coliphage T4 and the marine cyanophage S-PM2. Proc. Natl. Acad. Sci. USA 98:11411–11416.
46. Hennes, K. P., C. A. Suttle, and A. M. Chan. 1995. Fluorescently labeled virus probes show that natural virus populations can control the structure of marine microbial communities. Appl. Environ. Microbiol. 61:3623–3627.
47. Hewson, I., J. M. O'Neil, and W. C. Dennison. 2001. Virus-like particles associated with *Lyngbya majuscula* (Cyanophyta; Oscillatoriacea) bloom decline in Moreton Bay, Australia. Aquat. Microb. Ecol. 25:207–213.
48. Hwang-Lee, L., K. H. Blashka, J. Blamire, and R. E. McGowan. 1982. DNA metabolism during infection of *Anacystis nidulans* by cyanophage AS-1. 1. Identification of a unique species of DNA. Microbios 35:49–62.
49. Hwang-Lee, L., G. Cohn, L. Cosowsky, R. McGowan, and J. Blamire. 1985. DNA metabolism during infection of *Anacystis nidulans* by cyanophage AS-1. VII. UV-induced alterations of the AS-1/*A. nidulans* lytic cycle. Microbios 43:277–295.
50. Ishida, T., M. M. Watanabe, J. Sugiyama, and A. Yokota. 2001. Evidence for polyphyletic origin of the members of the orders of Oscillatoriales and Pleurocapsales as determined by 16S rDNA analysis. FEMS Microbiol. Lett. 201:79–82.
51. Kashyap, A. K., and S. Singh. 1989. Changes in photoelectron transport activity in cyanophage N-1- Infected cells of *Nostoc muscorum*. Curr. Microbiol. 18:151–155.
52. Khudyakov, I., and A. V. Matveev. 1982. New cyanophages S-4L and S-5L causing lysis of the cyanobacterium *Synechococcus elongatus*. Microbiology 51:91–97.
53. Khudyakov, I. Y., M. D. Kirnos, N. I. Alexandrushkina, and B. F. Vanyushin. 1978. Cyanophage S-2L contains DNA with 2,6-diaminopurine substituted for adenine. Virology 88:8–18.
54. Kirnos, M. D., I. Y. Khudyakov, N. I. Alexandrushkina, and B. F. Vanyushin. 1977. 2-Aminoadenine is an adenine substituting for a base in S-2L cyanophage DNA. Nature 270:369–370.
55. Koibong, L. I., L. Barksdale, and L. Garmise. 1961. Phenotypic alterations associated with the bacteriophage carrier state of *Shigella dysenteriae*. J. Gen. Microbiol. 24:355–367.
56. Koonin, E. V., K. S. Makarova, and L. Aravind. 2001. Horizontal gene transfer in prokaryotes: quantification and classification. Annu. Rev. Microbiol. 55:709–742.
57. Kroll, D., K. Meierhoff, N. Bechtold, M. Kinoshita, S. Westphal, U. C. Vothknecht, J. Soll, and P. Westhoff. 2001. VIPP1, a nuclear gene of *Arabidopsis thaliana* essential for thylakoid membrane formation. Proc. Natl. Acad. Sci. USA 98:4238–4242.
58. Lawrence, J. G., G. F. Hatfull, and R. W. Hendrix. 2002. Imbroglios of viral taxonomy: genetic exchange and failings of phenetic approaches. J. Bacteriol. 184:4891–4905.
59. Levine, E., and T. Thiel. 1987. UV-inducible DNA repair in the cyanobacteria *Anabaena* spp. J. Bacteriol. 169:3988–3993.
60. Li, W. K. W. 1995. Composition of ultraphytoplankton in the Central North-Atlantic. Mar. Ecol. Prog. Ser. 122:1–8.
61. Linnaeus, C. 1753. Species plantarum, exhibentes plantas rite cognitas, et genera relatas, cum differentus specifis, nominibus trivialibus, synonimis selectis, locis natalibus, secundum systema sexuale digestas II. Stockholm.
62. Liu, H. B., H. A. Nolla, and L. Campbell. 1997. *Prochlorococcus* growth rate and contribution to primary production in the equatorial and subtropical North Pacific Ocean. Aquat. Microb. Ecol. 12:39–47.
63. Lu, J., F. Chen, and R. E. Hodson. 2001. Distribution, isolation, host specificity, and diversity of cyanophages infecting marine *Synechococcus* spp. in river estuaries. Appl. Environ. Microbiol. 67:3285–3290.
64. Luftig, R., and R. Haselkorn. 1967. Morphology of a virus of blue-green algae and properties of its deoxyribonucleic acid. J. Virol. 1:344–361.
65. MacKenzie, J. J., and R. Haselkorn. 1972. Photosynthesis and the development of blue-green algal virus SM-1. Virology 49:517–521.
66. Mann, N. H. 2003. Phages of the marine cyanobacterial picophytoplankton. FEMS Microbiol. Rev. 27:17–34.
67. Mann, N. H., A. Cook, A. Millard, S. Bailey, and M. Clokie. 2003. Marine ecosystems: bacterial photosynthesis genes in a virus. Nature 424:741.
68. Marston, M. F., and J. L. Sallee. 2003. Genetic diversity and temporal variation in the cyanophage community infecting marine *Synechococcus* species in Rhode Island's coastal waters. Appl. Environ. Microbiol. 69:4639–4647.
69. Martin, E., and R. Benson. 1988. Phages of cyanobacteria, pp. 607. *In* R. Calendar (ed.) The Bacteriophages, vol. 2. Plenum Press, New York.

70. McDaniel, L., L. A. Houchin, S. J. Williamson, and J. H. Paul. 2002. Plankton blooms: lysogeny in marine *Synechococcus*. Nature 415:496.
71. Miller, R. V. 2001. Environmental bacteriophage–host interactions: factors contribution to natural transduction. Antonie van Leeuwenhoek 79:141–147.
72. Moisa, I., E. Sotropa, and V. Velehorsch. 1981. Investigations on the presence of cyanophages in fresh and sea waters of Romania. Rev. Roum. Med. Virol. 32:127–132.
73. Muradov, M. M., F. D. Kamilova, G. V. Cherkasova, R. S. Mukhamedov, A. A. Abdukarimov, and A. G. Khalmuradov. 1990. Detection of NP-1T sequences homologous to the *Klebsiella pneumoniae* structural genes for nitrogenase (*nif H, D, K*) in the DNA of different strains of temperate cyanophages. Mol. Biol. 24:1022–1026.
74. Nakajima, Y., S. Yoshida, Y. Inoue, and T. Ono. 1996. Occupation of the Q(B)-binding pocket by a photosystem II inhibitor triggers dark cleavage of the D1 protein subjected to brief preillumination. J. Biol. Chem. 271:17383–17389.
75. Ohki, K. 1999. A possible role of temperate phage in the regulation of trichodesmium biomass. Bull. Inst. Oceanogr. Monaco 19:287–291.
76. Ohki, K., and Y. Fujita. 1996. Occurrence of a temperate cyanophage lysogenizing the marine cyanophyte *Phormidium persicinum*. J. Phycol. 32:365–370.
77. Ortmann, A. C., J. E. Lawrence, and C. A. Suttle. 2002. Lysogeny and lytic viral production during a bloom of the cyanobacterium *Synechococcus* spp. Microb. Ecol. 43:225–231.
78. Padan, E., and M. Shilo. 1973. Cyanophages: viruses attacking blue-green algae. Bacteriol. Rev. 37:343–370.
79. Padan, E., M. Shilo, and A. B. Oppenheim. 1972. Lysogeny of the blue-green alga *Plectonema boryanum* by LPP2-SPI cyanophage. Virology 47:525–526.
80. Palenik, B., B. Brahamsha, F. W. Larimer, M. Land, L. Hauser, P. Chain, J. Lamerdin, W. Regala, E. E. Allen, J. McCarren, I. Paulsen, A. Dufresne, F. Partensky, E. A. Webb, and J. Waterbury. 2003. The genome of a motile marine *Synechococcus*. Nature 424:1037–1042.
81. Palenik, B., and R. Haselkorn. 1992. Multiple evolutionary origins of prochlorophytes, the chlorophyll *b*-containing prokaryotes. Nature 355:265–267.
82. Proctor, L. M., and J. A. Fuhrman. 1990. Viral mortality of marine bacteria and cyanobacteria. Nature 343:60–62.
83. Rimon, A., and A. B. Oppenheim. 1975. Heat induction of the blue-green alga *Plectonema boryanum* lysogenic for the cyanophage SPlcts1. Virology 64:454–463.
84. Rimon, A., and A. B. Oppenheim. 1976. Protein synthesis following infection of the blue-green alga *Plectonema boryanum* with the temperature virus SPI and its ts mutants. Virology 71:444–452.
85. Rippka, R., J. Deruelles, J. B. Waterbury, M. Herdman, and R. Y. Stanier. 1979. Generic assignments, strain histories and properties of pure cultures of cyanobacteria. J. Gen. Microbiol. 111:1–61.
86. Rocap, G., F. W. Larimer, J. Lamerdin, S. Malfatti, P. Chain, N. A. Ahlgren, A. Arellano, M. Coleman, L. Hauser, W. R. Hess, Z. I. Johnson, M. Land, D. Lindell, A. F. Post, W. Regala, M. Shah, S. L. Shaw, C. Steglich, M. B. Sullivan, C. S. Ting, A. Tolonen, E. A. Webb, E. R. Zinser, and S. W. Chisholm. 2003. Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. Nature 424:1042–1047.
87. Rodda, K. M. 1996. Temporal and spatial dynamics of *Synechococcus* spp. and *Micromonas pusilla* host–viral systems. MA thesis, University of Texas at Austin.
88. Rohwer, F., and R. Edwards. 2002. The phage proteomic tree: a genome-based taxonomy for phage. J. Bacteriol. 184:4529–4535.
89. Safferman, R. S., R. E. Cannon, P. R. Desjardins, B. V. Gromov, R. Haselkorn, L. A. Sherman, and M. Shilo. 1983. Classification and nomenclature of viruses of cyanobacteria. Intervirology 19:61–66.
90. Safferman, R. S., and M. E. Morris. 1963. Algal virus: isolation. Science 140:679–680.
91. Safferman, R. S., I. R. Schneider, R. L. Steere, M. E. Morris, and T. O. Diener. 1969. Phycovirus SM-1: a virus infecting unicellular blue-green algae. Virology 37:386–395.
92. Samimi, B., and G. Drews. 1978. Adsorption of cyanophage AS-1 to unicellular cyanobacteria and isolation of receptor material from *Anacystis nidulans*. J. Virol. 25:164–174.
93. Samuel, A. D. T., J. D. Petersen, and T. S. Reese. 2001. Envelope structure of *Synechococcus* sp. WH8113, a nonflagellated swimming cyanobacterium. BMC Microbiol 1. [Online]. http://www.biomedcentral.com/1471–2180/1/4
94. Samuilov, V. D., G. Renger, V. Z. Paschenko, A. V. Oleskin, M. V. Gusev, O. N. Gubanova, S. S. Vasilev, and E. L. Barsky. 1995. Inhibition of photosynthetic oxygen evolution by protonophoric uncouplers. Photosynth. Res. 46:455–465.
95. Schmitz, O., N. F. Tsinoremas, M. R. Schaefer, S. Anandan, and S. S. Golden. 1999. General effect of photosynthetic electron transport inhibitors on translation precludes their use for investigating regulation of D1 biosynthesis in *Synechococcus* sp strain PCC 7942. Photosynth. Res. 62:261–271.
96. Sherman, L. A. 1976. Infection of *Synechococcus cedrorum* by the cyanophage AS-1M. III. Cellular metabolism and phage development. Virology 71:199–206.
97. Sherman, L. A., and R. M. J. Brown. 1978. Cyanophages and viruses of eukaryotic algae, pp. 145–234. *In* H. Fraenkel-Conrat and R. R. Wagner (eds.) Comprehensive Virology, vol. 12. Plenum Press, New York.
98. Sherman, L. A., and M. Connelly. 1976. Isolation and characterization of a cyanophage infecting the unicellular blue-green algae *A. nidulans* and *S. cedrorum*. Virology 71:540–554.
99. Sherman, L. A., and R. Haselkorn. 1971. Growth of the blue-green algae virus LPP-1 under conditions which impair photosynthesis. Virology 45:739–746.
100. Simon, L. D. 1969. The infection of *Escherichia coli* by T2 and T4 bacteriophages as seen in the electron microscope. III. Membrane-associated intracellular bacteriophages. Virology 38:285–298.
101. Singh, P. K. 1975. Photoreactivation of UV-irradiated blue-green algae and algal virus LPP-1. Arch. Microbiol. 103:297–302.

102. Singh, S., A. Bhatnagar, and A. K. Kashyap. 1994. Energetics of cyanophage N-1 multiplication in the diazotrophic cyanobacterium *Nostoc muscorum*. Microbios 78:259–265.

103. Sippola, K., and E. M. Aro. 2000. Expression of *psbA* genes is regulated at multiple levels in the cyanobacterium *Synechococcus* sp. PCC 7942. Photochem. Photobiol. 71:706–714.

104. Smarda, J., D. Smajs, J. Komrska, and V. Krzyzanek. 2002. S-layers on cell walls of cyanobacteria. Micron 33:257–277.

105. Smith, K. M., R. M. J. Brown, and P. L. Walne. 1966. Electron microscopy of the infection process of the blue-green alga virus. Virology 30:182–192.

106. Sode, K., R. Oonari, and M. Oozeki. 1997. Induction of a temperate marine cyanophage by heavy metal. J. Mar. Biotechnol. 5:178–180.

107. Sode, K., M. Oozeki, K. Asakawa, J. G. Burgess, and T. Matsunaga. 1994. Isolation of a marine cyanophage infecting the marine unicellular cyanobacterium, *Synechococcus* sp. NKBG 0429002. J. Mar. Biotechnol. 1:189–192.

108. Stanier, R. Y., W. R. Sistrom, T. A. Hansen, B. A. Whitton, R. W. Castenholz, N. Pfennig, V. N. Gorlenko, E. N. Kondratieva, K. E. Eimhjellen, R. Whittenbury, R. L. Gherna, and H. G. Trüper. 1978. Proposal to place the nomenclature of the cyanobacteria (blue-green algae) under the rules of the International Code of Nomenclature of Bacteria. Int. J. Syst. Bacteriol. 28:335–336.

109. Sullivan, M. B., J. B. Waterbury, and S. W. Chisholm. 2003. Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. Nature 424:1047–1051.

110. Suttle, C. A. 2000. Cyanophages and their role in the ecology of cyanobacteria, pp. 563–589. *In* B. A. Whitton and M. Potts (eds.) The Ecology of Cyanobacteria. Kluwer Academic, Dordrecht.

111. Suttle, C. A., and A. M. Chan. 1994. Dynamics and distribution of cyanophages and their effect on marine *Synechococcus* spp. Appl. Environ. Microbiol. 60:3167–3174.

112. Suttle, C. A., and A. M. Chan. 1993. Marine cyanophages infecting oceanic and coastal strains of *Synechococcus*: abundance, morphology, cross-infectivity and growth-characteristics. Mar. Ecol. Prog. Ser. 92:99–109.

113. Szekeres, M., A. E. Szmidt, and I. Torok. 1983. Evidence for a restriction/modification-like system in *Anacystis nidulans* infected by cyanophage AS-1. Eur. J. Biochem. 131:137–141.

114. Teklemariam, T. A., S. Demeter, Z. Deak, G. Suranyi, and G. Borbely. 1990. AS-1 cyanophage infection inhibits the photosynthetic electron flow of photosystem II in *Synechococcus* sp. PCC 6301, a cyanobacterium. FEBS Lett. 270:211–215.

115. Udvardy, J., B. Sivok, G. Borbely, and G. L. Farkas. 1976. Formation in the dark, of virus-induced deoxyribonuclease activity in *Anacystis nidulans*, an obligate photoautotroph. J. Bacteriol. 126:630–633.

116. Urbach, E., D. L. Robertson, and S. W. Chisholm. 1992. Multiple evolutionary origins of prochlorophytes within the cyanobacterial radiation. Nature 355:267–270.

117. van Hannen, E. J., G. Zwart, M. P. van Agterveld, H. J. Gons, J. Ebert, and H. J. Laanbroek. 1999. Changes in bacterial and eukaryotic community structure after mass lysis of filamentous cyanobacteria associated with viruses. Appl. Environ. Microbiol. 65:795–801.

118. Veldhuis, M. J. W., G. W. Kraay, J. D. L. VanBleijswijk, and M. A. Baars. 1997. Seasonal and spatial variability in phytoplankton biomass, productivity and growth in the northwestern Indian Ocean: the southwest and northeast monsoon, 1992–1993. Deep Sea Res. A 44:425–449.

119. Waterbury, J. B. 1999. The Cyanobacteria: isolation, purification, and identification. *In* M. Dworkin (ed.) The Prokaryotes: An Evolving Electronic Resource for the Microbiological Community, 3rd edition, release 3. Springer, New York. http://141.150.157.117:8080/prokPUB/index.htm

120. Waterbury, J. B., and R. Rippka. 1989. Subsection I. Order Chroococcales Wettstein 1924, Emend. Rippka et al., 1979. *In* J. T. Staley, M. P. Bryant, N. Pfennig, and J. G. Holt (eds.) Bergey's Manual of Systematic Bacteriology, vol. 3. Williams and Wilkins, Baltimore, Md.

121. Waterbury, J. B., and F. W. Valois. 1993. Resistance to co-occurring phages enables marine *Synechococcus* communities to coexist with cyanophages abundant in seawater. Appl. Environ. Microbiol. 59:3393–3399.

122. Weinbauer, M. G., S. W. Wilhelm, C. A. Suttle, R. J. Pledger, and D. L. Mitchell. 1999. Sunlight-induced DNA damage and resistance in natural viral communities. Aquat. Microb. Ecol. 17:111–120.

123. Westphal, S., L. Heins, J. Soll, and U. C. Vothknecht. 2001. *Vipp1* deletion mutant of *Synechocystis*: a connection between bacterial phage shock and thylakoid biogenesis? Proc. Natl. Acad. Sci. USA 98:4243–4248.

124. Whitton, B. A., and M. Potts. 2000. Introduction to the cyanobacteria, pp. 1–11. *In* B. A. Whitton and M. Potts (eds.) The Ecology of Cyanobacteria. Kluwer Academic, Dordrecht.

125. Wilmotte, A. 1994. Molecular evolution and taxonomy of the cyanobacteria, pp. 1–25. *In* D. A. Bryant (ed.) The Molecular Biology of Cyanobacteria. Kluwer Academic, Dordrecht.

126. Wilson, W. H., N. G. Carr, and N. H. Mann. 1996. The effect of phosphate status on the kinetics of cyanophage infection in the oceanic cyanobacterium *Synechococcus* sp WH7803. J. Phycol. 32:506–516.

127. Wilson, W. H., N. J. Fuller, I. R. Joint, and N. H. Mann. 1999. Analysis of cyanophage diversity and population structure in a south–north transect of the Atlantic Ocean. Bull. Inst. Oceanogr. Monaco 19:209–216.

128. Wilson, W. H., I. R. Joint, N. G. Carr, and N. H. Mann. 1993. Isolation and molecular characterization of 5 marine cyanophages propagated on *Synechococcus*

sp strain WH7803. Appl. Environ. Microbiol. 59:3736–3743.

129. Xu, X., I. Khudyakov, and C. P. Wolk. 1997. Lipopolysaccharide dependence of cyanophage sensitivity and aerobic nitrogen fixation in *Anabaena* sp. strain PCC 7120. J. Bacteriol. 179:2884–2891.

130. Zhong, Y., F. Chen, S. W. Wilhelm, L. Poorvin, and R. E. Hodson. 2002. Phylogenetic diversity of marine cyanophage isolates and natural virus communities as revealed by sequences of viral capsid assembly protein gene g20. Appl. Environ. Microbiol. 68:1576–1584.

33

Marine Phages

ROBERT V. MILLER

Two decades ago marine bacteriophages were unimportant to microbial ecologists. After all, bacteriophages could only be important in certain environments where their concentrations were high enough to have an effect on bacterial populations. In most microbiologists' minds, this limited them to just a few ecosystems such as waste treatment facilities, cheese and yogurt production facilities, and the microbiology laboratory. It certainly did not include marine environments! Of course, specific phages such as PM2 were of interest (18, 87; see also chapter 14) because of their cytology and molecular microbiology. The few studies that had addressed the frequency of viruses in the oceans, however, showed that marine bacteriophages were not prevalent enough to affect marine ecosystems.

Two important developments in the 1980s changed this picture and stimulated interest in aquatic bacteriophages. The first arose from concerns originating from the newly emerging environmental biotechnology industry's use of genetically engineered microorganisms. Many feared that the recombinant sequences would escape to naturally occurring bacteria by horizontal gene exchange (36). Soon it was demonstrated that transduction was a viable mechanism for genetic exchange in aquatic environments (27, 37, 47, 68, 70, 72, 73). Still, phages were considered to be too infrequent in the aquatic environment to make this a real possibility. The second development, however, eliminated this objection and clearly demonstrated the importance of bacteriophages in marine environments. In 1989, Bratbak, Heldal and others (5, 8, 11) demonstrated that in many aquatic environments bacteriophages were present in very high concentrations and that they often exceeded by one to two orders of magnitude the concentrations of bacterioplankton that were their hosts. This report was quickly followed by several confirmatory publications (42, 64) that made it clear bacteriophages are indeed real players in the ecology of marine habitats.

This chapter is designed to provide the reader with an up-to-date and informed overview of the current understanding of the abundance of bacteriophages in aquatic, especially marine, environments. It is meant to demonstrate that bacterial viruses are important members of these ecosystems. It will explore the current controversy over the role of bacteriophages in the microbial loops of marine food webs (12). More detailed accounts of the subject can be found in several recent reviews (10, 20, 63, 79, 84, 100). See also chapter 32, which reviews cyanophages, the viruses of cyanobacteria.

The Abundance of Bacteriophages in Marine Environments

Why the Numbers Game?

Because bacteriophages are intracellular parasites of bacteria, they must find and infect a host bacterium in order to propagate. As the infective particle is biologically inert, virions encounter potential host organisms by simply diffusing through the suspending medium until they collide with a bacterium. If this collision results in the phage virion becoming attached to a host cell receptor, then infection follows (41, 76).

Since these collisions appear to occur at random, adsorption follows first-order kinetics (30, 76). As a consequence, in laboratory studies at least, the kinetics of phage attachment is generally found to be dependent on the concentration of both bacterial hosts and bacteriophages (32). Since most bacterial cells have the capacity to adsorb many bacteriophage virions before their receptors are saturated, the distribution of infected and uninfected bacteria in a population follows a Poisson distribution where the fraction of uninfected hosts (B_u) is given by the formula $B_u = Be^{-(PBR)}$, where B is the total concentration of bacteria per cubic centimeter and PBR is the phage-to-bacterium ratio found in the environment (32, 41). When the PBR is low (<0.1), B_u is very large, and the number of infected cells is an insignificant fraction of the total population (41).

Until the observations of Bratbak, Heldal and others (5, 8, 11) in 1989 and 1990, it was assumed that the number of marine bacteriophages was very low and, therefore, that the PBR in these environments would be <0.1 (41). Earlier studies had depended on enumeration of virions by observation of the consequences of virulent infection of a specific strain of host organism (44), and that led to gross underestimations of virion densities. For instance, in 1960 Spencer (75) detected four phage strains in North Sea water. They ranged in concentration from 0.1 to 10 plaque-forming units (PFU)/ml. A 1987 report by Moebus (44) found 1–3 PFU/ml for phages infecting each of five unidentified marine bacterial isolates. The only report of substantial numbers of phages was in the brackish Kiel Bight of the Baltic Sea where Ahrens (2) found phages of *Agrobacterium stellatum* in numbers as high as 10^4 PFU/ml. In these same environments, the numbers of bacteria were found to be significantly higher, with cell counts often as high as 10^4–10^6 CFU/ml (41). Hence, PBRs in the marine environment appeared to be very low, and it was reasonable to assume that phages could not have a significant influence on the make-up or density of bacterial populations, mediate gene transfer, or affect the food chain of the ecosystem in any significant way (41).

Current Estimates of Bacterial Virus and Virus-like Particles in Marine Environments

The 1990s brought new techniques including transmission electron microscopy (TEM; 9) to the isolation and enumeration of marine bacteriophages (39, 78), and it soon became clear that the true numbers of phage-like particles often exceeded 10^8/ml (41). Combined with other modern techniques including new methods of concentration (31, 58), epifluorescence (22, 51), flow cytometry (35), and pulse-field gel electrophoresis (17),TEM (33) allowed a truer picture of the high concentration of bacteriophages in aquatic environments. (A complete discussion of these techniques including their strengths and shortcomings can be found in Wommack and Colwell (100).)

The earliest report using TEM was actually in 1979, when Torrella and Morita (85) reported the numbers of bacteriophage particles in Yaquina Bay, Oregon to be as high as 10^3 particles/ml in areas where low concentrations of dissolved organics where observed and as high as 10^4 particles/ml where concentrations of organic material were high. The authors were careful to point out that no enrichments for bacteriophages were made and that they collected particles by employing a 0.2 μm pore-size filter. Therefore, they suspected that their numbers represented minimum estimates of phage concentrations at best. While Torrella and Morita demonstrated higher concentrations of marine bacteriophages than had been observed in other studies, their work may not have made the impact of later studies because no comparison with the number of bacterial hosts was made (i.e., no PBRs were estimated).

This was rectified in 1989 when Bergh and coworkers (5) demonstrated not only high concentrations of bacteriophages in several marine environments but that the PBRs in these environments were also high (>1.0). Phage concentrations of 2×10^8/ml were reported in environments that supported bacterial concentrations of only 6×10^6/ml (PBR $=33$). Assuming that this environment contained a minimum of 100 different phage–host systems, the authors estimated that the rate of phage adsorption might be as high as 2.5 particles/min per milliliter (5).

Bergh et al.'s (5) report was quickly followed by others (5, 8, 11, 42, 64, 82) which confirmed their findings and demonstrated that bacteriophage concentrations of this magnitude were common. Many of these studies related the numbers of bacteriophages to the numbers of potential hosts in these environments, allowing the calculation of PBRs for those environments (table 33.1). Phages could no longer be neglected in describing aquatic environment.

Suttle et al. (82) used TEM to show that natural marine water contains between 10^6 and 10^9 virus particles per milliliter. They demonstrated that these particles included viruses that infect a variety of microorganisms, in addition to bacteria, including diatoms, cryptophytes, prasinophytes, and cyanobacteria and that viral particle concentrations were higher in estuarine waters than in the open ocean. Thus, marine viruses became known as important effectors of populations of all types of microorganisms in marine environments.

Paul et al. (59, 60) studied bacteriophage concentrations at Key Largo, Florida (659), and at Mamala Bay, Oahu, Hawaii (60). In both environments they found that bacteriophage counts were highest in eutrophic areas (approximately 10^7/ml) and declined to around 10^6 particles/ml in offshore waters. Interestingly, the concentration of phages was inversely proportional to water salinity (59).

Wichels et al. (95) investigated the diversity of phages in a collection of 85 phages isolated from the North Sea. They found that the majority were members of the family *Myoviridae*, but members of the families *Siphoviridae* and *Podoviridae* were also in the collection. They determined that the phages belonged to 13 different species and that they all infected Gram-negative, facultatively anaerobic, motile, coccoid hosts that appeared to belong to the γ subdivision of the *Proteobacteria*. Thus, the potential of viruses to control bacterial diversity as well as numbers in natural marine environments becomes apparent.

Spatial and Temporal Variation in Bacteriophage Concentrations and PBRs

The size of bacteriophage populations has often been observed to vary seasonally. Bergh et al.'s (5) 1989 paper was the first to observe seasonal variation in the number of bacteriophages. Several orders of magnitude separated high summer frequencies from low winter numbers. In a subsequent report, Bratbak et al. (11) found that the

Table 33-1 Phage-to-Bacterium Ratios (PBR) and Concentrations of Virus-Like Particles in Various Marine Habitats

Marine environment	Virus-like particles ($\times 10^6$/ml)	PBR	Reference
Estuarine			
Chesapeake Bay, USA	3–140	3–26	101
Chesapeake Bay, USA	10	3	5
Lake Saelenvannet, Norway	20–300	20–80	89
Tampa Bay, USA	5–16	0.4–9	15
Tampa Bay, USA	5–20	0.9–9	25
Cosatal ocean			
Arctic Ocean, Resolute, Canada (sea ice)	9–430	10–72	34
Arctic Ocean, Resolute, Canada (seawater)	1.1	10–20	34
Bering and Chukchi Seas	2.5–36	5–5	77
North Adriatic Sea	9–130	5–25	93
Pacific Ocean, Japan	1–40	2–18	22
Paradise Harbor, Antarctica	0.2–1.3	0.7–6	7
Raunefjorden, Norway	0.01–10	<1–36	5
Southern California Bight, Santa Monica, USA	18	14	51
Open ocean			
Barents Sea	0.06	3	5
North Atlantic	14	50	5
North Pacific (subarctic)	0.06–0.4	1–4.5	23
North Pacific (subtropical)	0.4–2	1–9	23

concentration of phages rose from 5×10^5/ml in early spring to a maximum of more than 10^7/ml during a late summer–fall diatom bloom in a Norwegian fjord. Wommack et al. (101) found that the numbers of bacteriophages in Chesapeake Bay showed a peak from August to October with phage counts ranging from 10^6 to 10^8 particles/ml. Virus counts were always at least 3 times greater than bacterial counts and PBRs as high as 25 were reported at some samplings.

Weinbauer *et al.* (93) studied diel, seasonal, and depth-related variation in viral concentrations in the Northern Adriatic Sea. They found that during periods of water stratification, the highest numbers of phages and the highest PBR ratios were found at the thermocline. They speculated that this was due to the higher microbial biomass found there. In their studies, phage concentrations showed a strong seasonal variation, with phage abundance greatest in the fall (10^8–10^9 particles/ml) and lower in the winter ($<10^7$/ml). PBRs averaged 15 throughout the year, clearly demonstrating the potential of viruses to be key players in the ecology of the Adriatic.

Seasonal changes in the frequency of lysogenized bacteria were observed by Cochran and Paul (15). They found that samples that displayed prophage induction were plentiful during the warmer months, but no induction was observed in November, December, and January. Frequencies of inducible lysogens in the bacterial population ranged from undetectable in the winter months to as high as 37% in October, although the average varied around 10%.

Lytic Infection, Lysogeny, and Pseudolysogeny: Alternate Life-styles for the Wet and Small

Several studies have demonstrated the short infective life of bacteriophage virions in aquatic environments (73, 100). Decay rates of 5–30% per hour are not uncommon. This has led several investigators to explore the ways in which viruses maintain themselves in these environments. Aquatic environments are characterized by slow-growing bacterial hosts maintained at relative low concentrations of $<10^5$/ml (10, 11, 84). These are the very conditions that favor a temperate life-style (42, 100) and there have been many temperate bacteriophages (including temperate cyanophages; 54) isolated from marine environments (4, 29, 43, 48, 54, 56, 66, 99). In fact, Ackerman and DuBow (1), following a thorough investigation of the literature, estimated that, while the frequency of lysogeny varies among different bacterial taxonomic groups, between 21% and 60% of environmental bacteria are lysogens.

Since free virions of temperate and lytic bacteriophages cannot be distinguished via microscopy, investigators have turned to a variety of chemical and physical treatments including in situ hybridization (52, 53) to identify environmental lysogens. Jiang and Paul (26) used mitomycin C induction to estimate that 38% of the bacterial population in estuarine environments is lysogenized. Detectable lysogens were less frequent offshore.

In a later study, Jiang and Paul (28) explored the frequency of lysogens among 116 bacterial isolates from various marine environments. More than 40% of the strains contained inducible prophage. A higher percentage of lysogenized bacteria was found among isolates from oligotrophic environments than from coastal or estuarine waters. These observations are consistent with the assumption that lysogeny will be a more important life-style among bacteriophages in environments where hosts are few and energy is limited.

Weinbauer and Suttle (91, 92) found the frequency of inducible lysogens to be low (<5%) in bacterial populations in the Gulf of Mexico. The frequency was higher in offshore (2–11%) populations than it was in coastal water (1–2%). While the total numbers of inducible lysogens were lower than in some other studies, these data are also consistent with the assumption that lysogeny is more important in oligotrophic environments.

In addition to lytic and temperate growth, there is mounting evidence for an intermediate state, pseudolysogeny, of bacterium–bacteriophage interaction in natural environments. Simply stated, pseudolysogeny is a phage carrier state resembling lysogeny in that after infection bacteriophage can either enter a cryptic, intercellular state or sustain rapid lytic infection. However, unlike true lysogeny, pseudolysogeny does not involve integration of host and phage genomic DNA and the phage DNA is neither replicated nor segregated equally into all progeny cells (1, 69).

Many reports of pseudolysogeny (45, 46, 52, 67, 97) or of phage–host systems with characteristics of pseudolysogeny (99) in aquatic environments have appeared in the literature (for a review see 40). The bacteriophage Hs 1 of *Halobacterium salinarium* serves to illustrate many of the characteristics of pseudolysogeny and was one of the first marine archael systems to be studied (67, 86). In culture, infections of Hs 1 were characterized by sporadic lysis of major portions of infected batch cultures, yet it proved impossible to subculture stable lysogenic clones from the survivors. The interaction between host and phage changed with the concentration of salt in the medium. At 17.5% [wt/vol] NaCl (the lower limit for host survival) phage infection appeared to be virulent. However, as the concentration of salt approached 30%, the majority of infections led to nonproductive phage carrier states (pseudolysogens). Such conditions favor phage survival in environments where host growth is not favored (86). These observations demonstrate an essential characteristic of pseudolysogeny: phage production is regulated by environmental conditions that dictate host growth and survival (69, 100).

Following careful scrutiny of numerous marine phage–host relationships, Moebus (44, 45) has concluded that pseudolysogeny is a common phenomenon in marine ecosystems. Ackerman and DuBow (1) extended this generalization to include bacteriophage–host systems in general. Wommack and Colwell (100) suggest that pseudolysogeny affords "phage populations a means of quickly reacting to environmental changes." Ripp and Miller (40, 69) suggest that it affords environmental phages a mechanism with which to extend their infective half-life in environments with limited nutrients and capacity for viral growth until nutrients are available. Thus, Wommack and Colwell (100) believe that influxes of limited amounts of nutrients into a nutrient-limited marine ecosystem can stimulate both bacterial production and bacterial mortality through lytic activation of preprophage (the carrier state of the viral genome within the pseudolysogenized host envisioned by Ripp and Miller; 69).

Factors Affecting Bacteriophage Concentration

We have already seen that many environmental factors such as time of year (5, 15, 93, 100, 101), concentration of dissolved organics (59, 60, 82, 100), and salinity (59, 62, 100, 104) influence bacteriophage abundance and PBRs in marine environments. Other studies have explored the effects of other environmental components on phage populations.

Babich and Stotzky (3) investigated the differential toxicities of mercury to bacteria and bacteriophages in aquatic environments. They found that chloride salts of mercury were less toxic that was metallic mercury. This correlated with the fact that mercury was not as toxic to bacteria and bacteriophages in high-chlorine-containing marine waters than it was to bacteria and phages in fresh water.

A primary effector of bacteriophage virion stability in marine environments is solar ultraviolet (UV) light. Noble and Fuhrman (50) found that decay rates of several viruses in coastal waters were almost twice as fast in full sunlight as in the dark. Wommack et al. (102) tested decay rates of two *Aeromonas* phages. They incubated microcosms containing their phage–host test systems in Chesapeake Bay and in the York River estuary. Three types of microcosms were used. The first group was surface-incubated and unshielded (full sunlight). The second group was surface-incubated but covered (dark). The third group was unshielded but incubated at a depth of 1 m. Decay rates of phage infectivity were double in the first group compared with either of the other groups. These latter two groups showed decay rates equal to values obtained in the laboratory. However, the rates of destruction of phage particles were similar under all conditions and slower than any of the decay rates. Destruction of particles appears to be a process separate from loss of infectivity. This observation sounds a cautionary note for those studies that have used only TEM observation to determine the number of phages active in marine environments.

Suttle's group, currently at the University of British Columbia, has carried out an in-depth study to determine the effects of UV damage and repair on marine bacteriophages (80, 81, 94). They found that not only were UVB (290–320 nm) wavelengths damaging, but UVA

(320–300 nm) and photosynthetic light (400–700 nm) could also reduce phage infectivity (80). They studied host-associated repair of *Vibrio natriegens* bacteriophage conducted in offshore, coastal, and estuarine waters (94). In these experiments, light-dependent repair (probably photoreactivation using 370–550 nm light) compensated for a large fraction of the sunlight-induced damage to phage DNA (94). Photoreactivation appears to be essential to maintaining high concentrations of viable viruses in surface marine waters.

Control of Bacterial Population Densities and Implications for the Marine Food Web

Phage-Induced Mortality

In one of the earliest studies using TEM, Proctor and Fuhrman (64) demonstrated that a significant fraction of bacterial mortality in the ocean was due to viral infection. They observed that up to 7% of the heterotrophic bacteria and 5% of the cyanobacteria from diverse marine locations contained mature phages. These data suggested that up to 70% of the prokaryotes in these environments were infected with phages at any given time and that up to 30% of cyanobacterial and 60% of heterotrophic bacterial mortality was due to bacteriophage infection.

In a later, more detailed study, Proctor and Fuhrman (65) modified their expectations somewhat, but still estimated that mortality due to bacteriophage-induced lysis ranged from a low of 3% to a high of 62% for free-living bacteria and 52% for particle-associated cells. These studies warned that bacteriophage infection exerts a significant influence on carbon and nitrogen cycling in marine food webs.

Weinbauer and Suttle (91, 92) found that even in populations that contained low frequencies of inducible prophages (2–11% of the population), as much as 5% of the total bacterial mortality could be accounted for by induction of lysogens to lytic growth. Cochran et al. (16) demonstrated that many environmentally important pollutants, in addition to UV light, can act as inducing agents for natural lysogens in the Gulf of Mexico.

Miller (38) prepared microcosms of sterilized water from the Gulf of Mexico and inoculated them with a lysogen of *Vibrio parahaemolyticus.* These microcosms were incubated for a 3 day period on the deck of a research vessel during a research cruise of the Gulf. One half of the microcosms were exposed to natural solar radiation (containing UV light); the other half were covered to protect them from solar radiation. Results indicated that mortality of the lysogenic bacteria was dramatically increased in microcosms exposed to sunlight. Likewise, the number of virions isolated from the microcosms was much higher in sunlight-exposed chambers. In this time of increased solar UV exposure, due to ozone thinning, increased bacterial mortality due to induction of naturally occurring lysogens must be considered in estimating the effects of stratospheric ozone depletion on marine food webs.

Tuomi *et al.* (88) found that increased availability of carbon and energy in a seawater microbial community stimulated virus production to a greater extent than it stimulated bacterial biomass production. They found that significant increases in the PBR were brought about by these conditions. Although not investigated by these authors, these data suggest that the addition of nutrients to the environment leads to the activation of lysogens or pseudolysogens to the production of phage virions. These data are consistent with the finding of Ripp and Miller (69) that addition of an energy source to a freshwater microcosm led pseudolysogens of *Pseudomonas aeruginosa* starved for 15 days to lyse, releasing phage particles. It is well established that nutrient status influences the decision between lytic and temperate growth upon primary infection (98). The data of Tuomi et al. (88) suggest that a close link between nutrient concentration and initiation of phage production also exists in aquatic environments.

Control Microbial Population Size

The realization that bacteriophage concentrations were high enough to affect bacterial population size and rates of mortality turned the accepted dogma concerning food webs in the ocean upside down. Significant bacterial lysis by phages would alter the amount of dissolved carbon and the movement of carbon and energy through the food chain from bacteria to grazers and upwards to higher organisms. In addition to bacteriophages, the concentrations of viruses of other microorganisms, including primary producers (algae and cyanobacteria), were high enough to alter the accepted view of the contributions of algae and cyanobacteria to the movement of carbon through the marine food web.

Wommack and Colwell (100) developed a conceptual model of the microbial loop of the marine food web that includes viruses and viral lysis (figure 33-1). Similar ideas were voiced earlier by Fuhrman (19). These models demonstrate that bacteriophage activity enhances the flux of bacterial biomass into the dissolved organic material (DOM) pool. In addition, viral lysis of photosynthetic algae and cyanobacteria is found to augment the flux of photosynthetically fixed carbon into the DOM pool. Thus, the effect of viral lysis is to divert carbon away from mesozooplankton consumers (grazers) and into the DOM pool (19, 96, 100). Viral lysis is extremely efficient in moving biomass into the DOM pool because essentially all of the cell's contents move directly from the bacterial cell into the DOM pool (19, 100).

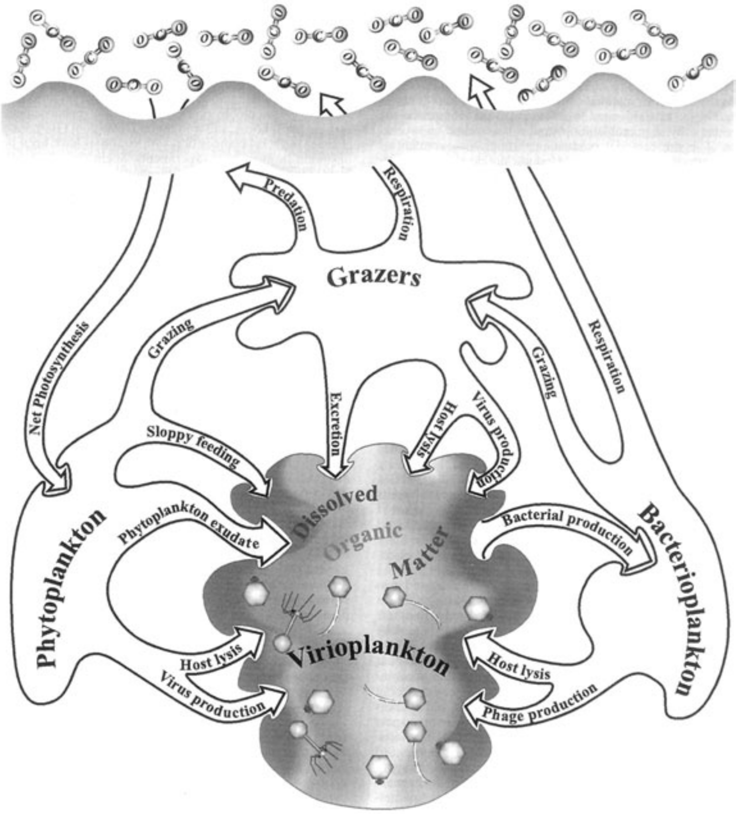

Predation
Respiration
Grazers
Net Photosynthesis
Grazing
Excretion
Host lysis
Virus production
Grazing
Respiration
Sloppy feeding
Dissolved
Organic
Matter
Phytoplankton
Phytoplankton exudate
Bacterial production
Bacterioplankton
Virioplankton
Host lysis
Virus production
Host lysis
Phage production

Figure 33-1 Microbial viruses and the microbial loop. A schematic diagram highlighting the potential role of virus infection and virus-induced cell lysis in the production of DOM in aquatic ecosystems. Reproduced with permission from Wommack and Colwell (100).

Bratbak et al. (11) hypothesized that phage lysis of bacterioplankton was a source of nutrient-rich growth substrate for bacterial populations. By allowing rapid recycling of carbon between bacterial biomass and DOM, phage infection allows bacterial populations to be sustained at higher levels than would be possible if no bacteriophage activity were present. This predicts that the results of active bacteriophage lysis are (i) increases in bacterial biomass and (ii) less transfer of organic matter to higher trophic levels.

Fuhrman (19) compared two hypothetical marine food webs. The first allowed for viral lysis of host cells; the second did not. If viral lysis was kept at a level of 50% of total bacterial mortality in the first system, there was a 27% increase in bacterial production rates and a 37% decrease in export of DOM to nanozooplankton grazers compared with the web where no bacteriophage lysis was allowed. This resulted in a net loss of 25% in nanozooplankton production. A later expansion of the model by Fuhrman and Suttle (21) included both viral infection of phytoplankton and loss of viroplankton to consumption by nanozooplankton. Even in this model, similar increases in bacterial production and reductions in nanozooplankton production were predicted.

Murray and Eldridge (49) carried out intensive theoretical exploration of the impact of viruses on aquatic microbial food webs. They allowed growth efficiency, recycling efficiency, and virus-induced mortality to vary under both oligotrophic and mesotrophic nutrient regimes and found that bacteriophage infection had the greatest impact in oligotrophic environments where recycling of organic matter predominates. Smaller effects were predicted in mesotrophic environments.

Bacteriophage Effects on Microbial Diveristy and Evolution

Following a careful analysis of all available literature, Wommack and Colwell concluded that approximately "20% or less of bacterioplankton and phytoplankton mortality is attributable to viral infection. Thus, viruses play a modest [but appreciable] role in controlling the population densities of bacteria" (100). Phages may have an even greater influence on species diversity for two reasons. First, bacteriophages can increase diversity by, as Thingstad (83) has speculated, "killing the winner" and allowing species to survive that would otherwise be eliminated by the dominant species. Second, they can increase diversity at the genetic level by mediating genetic exchange among bacterial strains and even species via transduction (68, 100).

Species Diversity

Recently, Wommack and Colwell (100) modeled bacteriophage influence on host community diversity. Their model predicted blooms of host bacteria due to changes in nutrients or environmental factors followed by killing of hosts and short-lived blooms of bacteriophages. Such blooms have been observed in situ (5, 55, 89, 105).

Hennes et al. (24) monitored population densities of the *Vibrio natriegens* strain PWH3a and its fluorescently labeled phage PWH3a-P1 following their introduction into a seawater microcosm. PWH3a rapidly increased to 40% of the population and then declined to less than 2%. The titers of PWH3a-P1 phage, on the other hand, rose throughout the incubation period from undetectable levels to 10^8/ml (about 70% of the total viral population). These authors concluded that viruses present at low abundances in natural aquatic viral communities can control microbial community structure.

In a study conducted in Chesapeake Bay, Wommack et al. (103) examined the presence of specific viruses using gene probes. They determined that titers of single viruses were highly localized within the Bay and changed over time with peaks and declines. The dynamics observed were consistent with models predicting bacteriophage control of blooms of single host strains (24, 100) that reduce their abundance and allow continued diversity in the population.

Thingstad (83) suggests that viral killing of dominant species ensures the coexistence of competing bacterial species. Bacteriophages act as a balancing force that allows bacterial species with different growth rates to coexist in a steady state due to the establishment of a "hierarchical theory of top-down control of diversity" (83). The limiting nutrient element that sets total biomass in the food web is at the top of this hierarchy. The second level is size-selective predation by grazers etc. that allows for distribution of the limiting element into different functional groups of the web. Finally, bacteriophage–host specificity allows further nutrient distribution into different species within a function group (83).

Genetic Diversity

Besides controlling diversity by regulating sizes of host populations, bacteriophages can increase diversity in bacterial gene pools through horizontal gene transfer by transduction. For instance, the acquisition of virulence determinants by disease-causing *Vibrio cholerae* is associated with transduction and lysogenic conversion (6, 13, 90). Until the realization that there are high concentrations of bacteriophages in the aquatic environment, transduction was dismissed as a viable gene transfer system for several reasons. First, most bacteriophages are restricted to infection of a narrow range of bacterial hosts (41), and transduction is by its nature a reductive process that kills the genetic donor in the process of producing transducing particles (36). However, on the positive side, transduction may be favorable to conjugation and transformation because the gene-transfer elements (transducing particles) are relatively long lived and transduction does not require cell-to-cell contact (62, 74). When the true concentrations of bacteriophage particles in the aquatic environment were realized, transduction suddenly became a viable option for gene transfer in these environments (37).

Transduction was demonstrated in the marine environment by Jiang and Paul (27) who established a model transduction system using marine phage–host isolates that complemented long-studied freshwater models (37, 38). They isolated a bacterium, D1B, from Mamala Bay, Hawaii that was identified as a *Flavobacterium* sp. (27), and a temperate phage, T-ϕD1B, that infected D1B. Transduction experiments using this phage–host system were carried out in microcosms in the Tampa Bay Estuary, and transduction rates as high as 10^{-7} transductants/PFU were observed for the transfer of the Tra^- plasmid pQSR50 (Kan^r, Str^r). Rates of 4×10^{-8} were documented in microcosms containing mixed bacterial communities from the Bay. Even though these rates are low, they would produce as many as 10^{14} transduction events each year in a marine environment the size of Tampa Bay (27).

Chiura (14) used virus-like particles released from isolates of various marine bacteria to transduce auxotrophic strains of *Escherichia coli* and *Bacillus subtilis* to prototrophy. They obtained transduction frequencies as high as 10^{-3}/particle. These studies demonstrated that naturally occurring virus-like particles could act as transducing agents.

The studies described above and a number of others have demonstrated the real potential for transduction in aquatic environments and several reviews on the subject have appeared (34, 36, 41, 57, 71). Clearly, transduction can no longer be dismissed as a phenomenon restricted to the microbiology laboratory and must be considered in assessing the evolutionary potential of bacteria in aquatic environments.

Conclusions

It is now well established that bacteriophages are an active and dynamic factor in marine ecosystems. Bacteriophages must be considered in any investigation of the importance of microorganisms to the ecology of all habitats found on our planet. Bacteriophages must be included in any models of natural food webs, and they must be considered in assessing the impact of pollutants and global environmental changes on the biosphere of the Earth.

References

1. Ackerman, H.W., and M. S. DuBow. 1987. Viruses of Prokaryotes: General Properties of Bacteriophages. CRC Press, Baca Raton, Fla.
2. Ahrens, R. 1971. Untersuchungen zur Verbreitung von Phagen der *Agrobacteriumi* in der Ostsee. Kieler Meeresforsch. 27:102–112.
3. Babich, H., and G. Stotzky. 1979. Differential toxicities of mercury to bacteria and bacteriophages in sea and in lake water. Can. J. Microbiol. 25:1252–1257.
4. Baross, J. A., J. Liston, and R. Y. Morita. 1978. Incidence of *Vibrio parahaemolyticus* bacteriophages and other *Vibrio* bacteriophages in marine samples. Appl. Environ. Microbiol. 36:492–499.
5. Bergh, O., K. Y. Borsheim, G. Bratbak, and M. Heldal. 1989. High abundance of viruses found in aquatic environments. Nature 340:467–468.
6. Bik, E. M., A. E. Bunshoten, R. D. Gouw, and F. R. Mooi. 1995. Genesis of the novel epidemic *Vibrio cholerae* O139 strain: evidence for horizontal transfer of genes involved in polysaccharide synthesis. EMBO J. 14:209–216.
7. Bird, D. F., and R. Maranger. 1993. Palmer LTER: aquatic virus abundances near the Antarctic Peninsula. Antarct. J. U. S. 28:234–235.
8. Borsheim, K. Y., G. Brantbak, and M. Heldal. 1990. Enumeration and biomass estimation of planktonic bacteria and viruses by transmission electron microscopy. Appl. Environ. Microbiol. 56:352–356.
9. Bratbak, G., and M. Heldal. 1993. Total count of viruses in aquatic environments, pp. 135–138. *In* P. F. Kemp, B.F. Sherr, E. B. Sherr, and J. J. Cole, (eds.) Handbook of Methods in Aquatic Microbial Ecology. Lewis Publishers, Boca Raton, Fla.
10. Bratbak, G., M. Heldal, A. Naess, and T. Poeggen. 1993. Viral impact on microbial communities, pp. 229–302. *In* R. Guerrero, and C. Pedros-Alio, (eds.) Trends in Microbial Ecology. Spanish Society for Microbiology, Barcelona.
11. Bratbak, G., M. Heldal, S. Norland, and T. F. Thingstad. 1990. Viruses as partners in spring bloom microbial trophodynamics. Appl. Environ. Microbiol. 56:1400–1405.
12. Bratbak, G., F. Thingstad, and M. Heldal. 1994. Viruses and the microbial loop. Microb. Ecol. 28:209–221.
13. Cheetham, B. F., and M. E. Katz. 1995. A role for bacteriophages in the evolution and transfer of bacterial virulence determinants. Mol. Microbiol. 18:201–208.
14. Chiura, H. X. 1997. Generalized gene transfer by virus-like particles from marine bacteria. Aquat. Microb. Ecol. 13:75–83.
15. Cochran, P. K., and J. H. Paul. 1998. Seasonal abundance of lysogenic bacteria in a subtropical estuary. Appl. Environ. Microbiol. 64:2308–2312.
16. Cochran, P. K., C. A. Kellogg, and J. H. Paul. 1998. Prophage induction of indigenous marine lysogenic bacteria by environmental pollutants. Mar. Ecol. Prog. Ser. 164:125–133.
17. Diez, B., J. Anton, N. Guixa-Boixereu, C. Pedros-Alio, and R. Rodriguez-Valera. 2000. Pulse-field gel electrophoresis analysis of virus assemblages present in a hypersaline environment. Int. Microbiol. 3:159–164.
18. Espejo, R. T., and E. S. Canelo. 1968. Properties of bacteriophage PM2: a lipid-containing bacterial virus. Virology 34:738–747.
19. Fuhrman, J. A. 1992. Bacterioplankton roles in cycling of organic matter: the microbial food web, pp. 361–383. *In* P. G. Falkowski, and A. D. Woodhead, (eds.) Primary Productivity and Biogeochemical Cycles in the Sea. Plenum Press, New York.
20. Fuhrman, J. A. 1999. Marine viruses and their biogeochemical and ecological effects. Nature 399:541–548.
21. Fuhrman, J. A., and C. A. Suttle. 1993. Viruses in marine planktonic systems. Oceanography 6:50–62.
22. Hara, S., K. Terauchi, and I. Koike. 1991. Abundance of viruses in marine water: assessment by epifluorescence and transmission electron microscopy. Appl. Environ. Microbiol. 57:2731–2734.
23. Hara, W., I. Koike, K. Terauchi, H. Kamiya, and E. Tanoue. 1996. Abundance of viruses in deep oceanic waters. Mar. Ecol. Prog. Ser. 145:269–277.
24. Hennes, K. P., C. A. Suttle, and A. M. Chan. 1995. Fluorescently labeled virus probes show that natural virus populations can control the structure of marine microbial communities. Appl. Environ. Microbiol. 61:3623–3627.
25. Jiang, S. C., and J. H. Paul. 1994. Seasonal and diel abundance of viruses and occurrence of lysogeny/bacteriocinogeny in the marine environment. Mar. Ecol. Prog. Ser. 104:163–172.
26. Jiang, S. C., and J. H. Paul. 1996. Occurrence of lysogenic bacteria in marine microbial communities as determined by prophage induction. Mar. Ecol. Prog. Ser. 142:27–38.
27. Jiang, S. C., and J. H. Paul. 1998. Gene transfer by transduction in the marine environment. Appl. Environ. Microbiol. 64:2780–2787.
28. Jiang, S. C., and J. H. Paul. 1998. Significance of lysogeny in the marine environment: studies with isolates and a model of lysogenic phage production. Microb. Ecol. 35:235–243.
29. Jiang, S. C., C. A. Kellogg, and J. H. Paul. 1998. Characterization of marine temperate phage–host systems isolated from Mamala Bay, Oahu, Hawaii. Appl. Environ. Microbiol. 64:535–542.
30. Kokjohn, T. A. 1989. Transduction: mechanism and potential for gene transfer in the environment, pp. 73–98. *In* S. B. Levy, and R. V. Miller, (eds.) Gene Transfer in the Environment. McGraw-Hill, New York.
31. Li, J.W., X.W. Wang, Q.Y. Rui, N. Song F. G. Zhang, Y. C. Ou, and F. H. Chao. 1998. A new and simple method for concentration of enteric viruses from water. J. Virol. Methods 74:99–108.
32. Luria, S. E., J. E. Darnell, Jr., D. Baltimore, and A. Campbell. 1978. General Virology, 3rd edn. Wiley, New York.
33. MacRae, J., and M. Srivastava. 1998. Detection of viruses by electron microscopy: an efficient approach. J. Virol. Methods 72:105–108.

34. Maranger, R., D. F. Bird, and S. K. Juniper. 1994. Viral and bacterial dynamics in Arctic sea ice during the spring algal bloom near Resolute, N.W.T., Canada. Mar. Ecol. Prog. Ser. 111:121–127.
35. Marie, D., C. P. D. Brussard, R. Thyrhaug, G. Bratbak, and D. Voulot. 1999. Enumeration of marine viruses in culture and natural samples by flow cytometry. Appl. Environ. Microbiol. 5:45–52.
36. Miller, R. V. 1988. Potential for transfer and establishment of engineered genetic sequences in the environment, pp. S23–S27. *In* J. Hodgson, and A. M. Sugden (eds.) Planned Release of Genetically Engineered Organisms (Trends Biotech. and Trends Ecol. Evolution Special Publication). Elsevier, Cambridge, UK.
37. Miller, R. V. 1998. Bacterial gene swapping in nature. Sci. Am. 278:66–71.
38. Miller, R. V. 2001. Environmental bacteriophage–host interactions: factors contributing to natural transduction. Antonie van Leeuwenhoek, 79:141–147.
39. Miller, R. V. 1998. Methods for enumeration and characterization of bacteriophages from environmental samples, pp. 218–235. *In* R. Burlage, (ed.) Techniques in Microbial Ecology. Oxford University Press, Oxford.
40. Miller, R.V., and S. A. Ripp. 2002. Pseudolysogeny: a bacteriophage strategy for increasing longevity in situ, pp. 81–94. *In* M. Syvanen, and C. Kado, (eds). Horizontal Gene Transfer, 2nd edn. Academic Press, San Diego.
41. Miller, R. V., and G. S. Sayler. 1992. Bacteriophage–host interactions in aquatic systems, pp. 176–193. *In* E. M. Wellington, and J. D. Van Elsas. (eds.) Genetic Interactions among Microorganisms in the Natural Environment. Pergamon Press, Oxford.
42. Miller, R. V., S. Ripp, J. Replicon, O. A. Ogunseitan, and T. A. Kokjohn. 1992. Virus-mediated gene transfer in freshwater environments, pp. 50–62. *In* M. J Gauthier, (ed.) Gene Transfers and Environment. Springer, Berlin.
43. Mitra, S. N., S. Kar, R. K. Ghosh, S. Pajni, and A. Ghosh. 1995. Presence of lysogenic phage in the outbreak strains of *Vibrio cholerae* O139. J. Med. Microbiol. 42:399–403.
44. Moebus, K. 1987. Ecology of marine bacteriophages, pp. 137–156. *In* S. M. Goyal, C. P. Gerba, and G. Bitton, (eds.) Phage Ecology. Wiley, New York.
45. Moebus, K. 1996. Marine bacteriophage reproduction under nutrient-limited growth of host bacteria. 2. Investigation with phage–host system [H3:H3/1]. Mar. Ecol. Prog. Ser. 144:13–22.
46. Moebus, K. 1997. Investigations of the marine lysogenic bacterium H24. 2. Development of pseudolysogeny in nutrient rich broth. Mar. Ecol. Prog. Ser. 148:229–240.
47. Morrison, D., R.V. Miller, and G. S. Sayler. 1978. Frequency of F116 mediated transduction of *Pseudomonas aeruginosa*: in natural freshwater environment. Appl. Environ. Microbiol. 36:724–730.
48. Muramatsu, K., and H. Matsumoto. 1991. Two generalized transducing phages in *Vibrio parahaemolyticus* and *Vibrio alginolyticus*. Microbiol. Immunol. 35:1073–1084.
49. Murray, A. G., and P. M. Eldridge. 1994. Marine viral ecology: incorporation of bacteriophage into the microbial planktonic food web paradigm. J. Plankton Res. 16:627–641.
50. Noble, R. T., and J. A. Fuhrman. 1997. Virus decay and its causes in coastal waters. Appl. Environ. Microibol. 63:77–83.
51. Noble, R. T., and J. A. Fuhrman. 1998. Use of SYBR Green I for rapid epifluorescence counts of marine viruses and bacteria. Aquat. Microb. Ecol. 14:113–118.
52. Ogunseitan, O. A., G. S. Sayler. and R. V. Miller. 1990. Dynamic interactions of *Pseudomonas aeruginosa* and bacteriophages in lake water. Microb. Ecol. 19:171–185.
53. Ogunseitan, O. A., G. S. Sayler and R. V. Miller. 1992. Application of DNA probes to the analysis of bacteriophage distribution patterns in the environment. Appl. Environ. Microbiol. 58:2046–2052.
54. Ohki, K., and Y. Fujita. 1996. Occurrence of a temperate cyanophage lysogenizing the marine cyanophyte *Phormidium persicinu*. J. Phycol. 32:365–370.
55. Padan, E., and M. Shilo. 1973. Cyanophages: viruses attacking blue-green algae. Bacteriol. Rev. 37:343–370.
56. Pajni, S., N. R. Chowdhury, A. Ghosh, S. Kar, and R. K. Ghosh. 1995. Characterization of phage phi O139, a *Vibrio cholerae* O139 temperate bacteriophage with cohesive DNA termini. FEMS Microbiol. Lett. 131:69–74.
57. Paul, J. H. 1999. Microbial gene transfer: an ecological perspective. J. Mol. Microbiol. Biotechnol. 1:45–50.
58. Paul, J. H., S. C. Jiang, and J. B. Rose. 1991. Concentration of viruses and dissolved DNA from aquatic environments by vortex flow filtration. Appl. Environ. Microbiol. 57:2197–2204.
59. Paul, J. H., J. B. Rose, S. C. Jiang, C. A. Kellogg, and I. Dickson. 1993. Distribution of viral abundance in the reef environment of Key Largo, Florida. Appl. Environ. Microbiol. 59:718–724.
60. Paul, J. H., J. B. Rose, S. C. Jiang, P. London, X. Xhou, and C. Kellogg. 1997. Coliphage and indigenous phage in Mamala Bay, Oahu, Hawaii. Appl. Environ. Microbiol. 63:133–138.
61. Pedros-Alio, C., J. K. Calderon-Paz, M. H. MacLean, G. Medina, C. Marrase, J. M. Gasol, and N. Guiza-Boixereu. 2000. The microbial food web along salinity gradients. FEMS Microb. Ecol. 32:143–155.
62. Pickup, R. W. 1992. Detection of gene transfer in aquatic environments, pp. 145–164. *In* E. M. H. Wellington, and J. D. van Elsas, (eds.) Genetic Interactions among Microorganisms in the Natural Environment. Pergamon Press, New York.
63. Proctor, L. M. 1997. Advances in the study of marine viruses. Microsc. Res. Tech. 37:136–161.
64. Proctor, L. M., and J. A. Fuhrman. 1990. Viral mortality of marine bacteria and cyanobacteria. Nature 343:60–62.
65. Proctor, L. M., A. Okubo, and J. A. Fuhrman. 1993. Calibrating estimates of phage-induced mortality in marine bacteria: ultrastructural studies of marine bacteriophage development from one-step growth experiments. Microb. Ecol. 25:161–182.
66. Rambler, M., and L. Margulis. 1979. An ultraviolet light induced bacteriophage in *Beneckea gazogenes*. Orig. Life 9:235–240.
67. Reiter, W. D., W. Zillig, and P. Palm. 1988. Archaebacterial viruses. Adv. Virus Res. 34:143–188.

68. Replicon, J., A. Frankfater, and R. V. Miller. 1995. A continuous culture model to examine factors that affect transduction among *Pseudomonas aeruginosa* strains in freshwater environments. Appl. Environ. Microbiol. 61:3359–3366.
69. Ripp, S., and R.V. Miller. 1997. The role of pseudolysogeny in bacteriophage–host interactions in a natural freshwater environment. Microbiology 143:2065–2070.
70. Ripp, S., O. A. Ogunseitan, and R. V. Miller. 1994. Transduction of a freshwater microbial community by a new *Pseudomonas aeruginosa* generalized transducing phage, UT1. Mol. Ecol. 3:121–126.
71. Saye, D. J., and R. V. Miller. 1989. The aquatic environment: consideration of horizontal gene transmission in a diversified habitat, pp. 223–259. *In* S. B. Levy, and R.V. Miller, (eds.) Gene Transfer in the Environment. McGraw-Hill, New York.
72. Saye, D. J., O. A. Ogunseitan, G. S. Sayler, and R. V. Miller. 1990. Transduction of linked chromosomal genes between *Pseudomonas aeruginosa* during incubation in situ in a freshwater habitat. Appl. Environ. Microbiol. 56:140–145.
73. Saye, D. J., O. Ogunseitan, G. S. Sayler, and R.V. Miller. 1987. Potential for transduction of plasmids in a natural freshwater environment: effect of plasmid donor concentration and a natural microbial community on transduction in *Pseudomonas aeruginosa*. Appl. Environ. Microbiol. 53:987–995.
74. Schicklmaier, P., and H. Schmieger. 1995. Frequency of generalized transducing phages in natural isolates of the *Salmonella typhimurium* complex. Appl. Environ. Microbiol. 61:1637–1640.
75. Spencer, R. 1960. Indigenous marine bacteriophages. J. Bacteriol. 79:614.
76. Stent, G. S. 1963. Molecular Biology of Bacterial Viruses. W. H. Freeman, San Francisco.
77. Steward, G. F., D. C. Smith, and F. Azam. 1996. Abundance and production of bacteria and viruses in the Bering and Chukchi seas. Mar. Ecol. Prog. Ser. 131:287–300.
78. Suttle, C. A. 1993. Enumeration and isolation of viruses, pp. 121–134. *In* P. F. Kemp, B. F. Sherr, E. F. Sherr, and J. J. Cole (eds.) Handbook of Methods in Aquatic Microbial Ecology. Lewis Publishers, Boca Raton, Fla.
79. Suttle, C. A. 1999. Do viruses control the oceans? Nat. Hist. 108:48–51.
80. Suttle, C. A., and A. M. Chan. 1992. Mechanisms and rates of decay of marine viruses in seawater. Appl. Environ. Microbiol. 58:3721–3729.
81. Suttle, C. A., and A. M. Chan. 1994. Dynamics and distribution of cyanophages and their effect on marine *Synechococcus* spp. Appl. Environ. Microbiol. 60:3167–3174.
82. Suttle, C. A., A. M. Chan, and M. T. Cottrell. 1990. Infection of phytoplankton by viruses and reduction of primary productivity. Nature 347:467–469.
83. Thingstad, T. F. 2000. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. Limnol. Oceanogr. 45:1320–1328.
84. Thingstad, T. F., M. Heldal, G. Bratbak, and I. Dundas. 1993. Are viruses important partners in pelagic food webs? Trends Ecol. Evol. 8:209–212.
85. Torrella, F., and R. Y. Morita. 1979. Evidence by electron micrographs for a high incidence of bacteriophage particles in the waters of Yaquina Bay, Oregon: ecological and taxonomical implications. Appl. Environ. Microbiol. 37:774–778.
86. Torsvik, T., and I. D. Dundas. 1980. Persisting phage infection in *Halobacterium salinarium* str. 1. J. Gen. Virol. 47:29–36.
87. Tsukagoshi, N., M. H. Petersen, U. Huber, R. M. Franklin, and J. Seelig. 1976. Phase transitions in the membrane of a marine bacterium, *Pseudomonas* BAL-31. Eur. J. Biochem. 62:257–262.
88. Tuomi, P., K. M. Fagerbakke, G. Bratbak, and M. Heldal. 1995. Nutritional enrichment of a microbial community: the effects on activity, elemental composition, and community structure and virus production. FEMS Microbiol. Ecol. 1:123–134.
89. Tuomi, P., T. Torsvik, M. Heldal, and G. Bratbak. 1997. Bacterial population dynamics in a meromictic lake. Appl. Environ. Microbiol. 63:2181–2188.
90. Waldor, M. K., and J. J. Mekalanos. 1996. Lysogenic conversion by a filamentous phage encoding cholera toxin. Science 272:1910–1914.
91. Weinbauer, M. G., and C. A. Suttle. 1996. Potential significance of lysogeny to bacteriophage production and bacterial mortality in coastal waters of the Gulf of Mexico. Appl. Environ. Microbiol. 62:4374–4380.
92. Weinbauer, M. G., and C. A. Suttle. 1999. Lysogeny and prophage induction in coastal and offshore bacterial communities. Aqua. Microb. Ecol. 18:217–225.
93. Weinbauer, M. G., D. Fuks, S. Puskaric, and P. Peduzzi. 1995. Diel, seasonal, and depth-related variability in viruses and dissolved DNA in the Northern Adriatic Sea. Microb. Ecol. 30:25–41.
94. Weinbauer, M. G., S. W. Wilhelm, C. A. Suttle and D. R. Carza. 1997. Photoreactivation compensates for UV damage and restores infectivity to natural marine virus communities. Appl. Environ. Microbiol. 63:2200–2205.
95. Wichels, A., S. S. Biel, H. R. Gelderblon, T. Brinkhoff, G. Muyzer, and C. Schutt. 1998. Bacteriophage diversity in the North Sea. Appl. Environ. Microbiol. 64:4128–4133.
96. Wilhelm, S.W., and C. A. Suttle. 1999. Viruses and nutrient cycles in the sea. Bioscience 49:781–788.
97. Williamson, S. J., M. R. McLaughlin, and J. H. Paul. 2001. Interaction of the ΦHSIC virus with its host: lysogeny or pseudolysogeny? Appl. Environ. Microbiol. 67:1682–1688.
98. Wilson, W. H., and N. H. Mann. 1997. Lysogenic and lytic viral production in marine microbial communities. Aquat. Microb. Ecol. 3:95–100.
99. Wilson, W. H., N. G. Carr, and N. H. Mann. 1996. The effect of phosphate status on the kinetics of cyanophage infection in the oceanic cyanobacterium *Synechococcus* sp. WH7803. J. Phycol. 32:506–516.
100. Wommack, K. E., and R. R. Colwell. 2000. virioplankton: viruses in aquatic ecosystems. Microbiol. Mol. Biol. Rev. 64:69–114.
101. Wommack, K. E., R. T. Hill, M. Kessel, E. Russek-Cohen, and R. R. Colwell. 1992. Distribution of viruses in Chesapeake Bay. Appl. Environ. Microbiol. 58:2965–2970.

102. Wommack, K. E., R. T. Hill, T. A. Muller, and R. R. Colwell. 1996. Effects of sunlight on bacteriophage viability and structure. Appl. Environ. Microbiol. 62:1336–1341.

103. Wommack, K. E., J. Ravel, R. T. Hill, and R. R. Colwell. 1999. Hybridization analysis of Chesapeake Bay virioplankton, Appl. Environ. Microbiol. 65:241–250.

104. Zachary, A. 1976. Physiology and ecology of bacteriophages of the marine bacterium *Beneckea natriegens*: salinity. Appl. Environ. Microbiol. 31:415–422.

105. Zingone, A. 1995. The role of viruses in the dynamics of phytoplankton blooms. G. Bot. Ital. 129:415–423.

34

Yersinia Phages

STEFAN HERTWIG
MIKAEL SKURNIK
BERND APPEL

The genus *Yersinia* contains 11 species, of which *Yersinia enterocolitica*, *Yersinia pestis*, and *Yersinia pseudotuberculosis* are pathogenic for humans and animals (5). While *Y. pestis* is the causative agent of plague, *Y. enterocolitica* and *Y. pseudotuberculosis* are mainly enteropathogenic. In comparison with the other pathogenic *Yersinia* species, *Y. enterocolitica* is very heterogeneous, comprising pathogenic and nonpathogenic strains which belong to about 70 serogroups. In order to differentiate *Yersinia* strains, a number of phages were isolated and used for typing. Several phage typing sets have been worked out for *Y. enterocolitica* (4, 30, 38). The sets contain temperate phages isolated from *Yersinia* strains (37, 39) or phages isolated from sewage (2, 8), whose origin is unknown. Three yersiniophages have been sequenced: phages ϕA1122, ϕYeO3-12, and PY54. This chapter is primarily an overview of the properties of those three phages.

Overview of Yersiniophage Characteristics

Host Range

Temperate *Y. enterocolitica* phages can be easily isolated since high frequencies (up to 86.4%) of lysogeny are observed (36, 58). In contrast, much lower frequencies have been determined for nonpathogenic *Yersinia* species (9) and there are only a few reports about the isolation of temperate phages from *Y. pestis* and *Y. pseudotuberculosis* (33, 40). Most of the temperate *Y. enterocolitica* phages showed a narrow host range and were not able to infect other *Yersinia* species, nor other *Enterobacteriaceae*, nor Gram-positive bacteria (23, 39). However, phages isolated from nonpathogenic *Yersinia* species are often active on pathogenic *Y. enterocolitica* strains (10, 12). Other than host ranges, derived from phage-typing experience, there is only scarce information available on *Yersinia* phages.

Morphologies

Kasatiya and Ackermann (23) studied the morphology of nine *Y. enterocolitica* phages and found four morphotypes. The phages had isometric or elongated heads and contractile or very short tails. Phages with isometric heads and contractile tails were also described by Kawaoka et al. (27). In a more comprehensive study, eight temperate yersiniophages were characterized on the basis of their morphology, host range, genome size, DNA homology, and protein composition (45). The phages contain double-stranded DNA with genome sizes of 40–60 kb. The morphology of the phages suggested that they belong to different families. However, DNA homologies were observed between members of the families *Myoviridae* and *Podoviridae*.

Temperature Dependence

Several authors noted that temperate yersiniophages are not able to be lytic at 37 °C (4, 39). Studies of growth-temperature-dependent variation of lipopolysaccharides indicated that the receptor for a number of yersiniophages is a glycoconjugate other than lipopolysaccharicharid, one that is synthesized at 25 °C but not at 37 °C (28, 29). Contrary to this report, Calvo et al. (11) showed that the receptor for a *Y. enterocolitica* O:3-specific phage was present at 37 °C and concluded that the phage was able to adsorb but not to replicate at this temperature.

Transduction

One of the phages in the study by Popp et al. (45), phage PY20 isolated from a pathogenic *Y. enterocolitica* O:3 strain, was used for transduction experiments. PY20 was able to transduce small *Yersinia* plasmids (4.3 and 5.8 kb) but not the 70 kb *Yersinia* virulence plasmid, pYV. The same result was obtained with phage mixtures isolated from sewage that had been tested for their ability to infect *Yersinia* strains

(18). Obviously these phages did not have the packaging capacity to encapsidate the relatively large pYV plasmid. In contrast, phage P1 was shown to transfer the pYV *Yersinia* virulence plasmid of *Y. pestis* to *Y. pseudotuberculosis* (59).

Sequenced Yersiniophages

Up to now the genomes of three yersiniophages (ϕA1122, ϕYeO3-12, and PY54) have been sequenced. The phages ϕYeO3-12 and PY54 are discussed in the following sections.

Yersiniophage ϕA1122

Overview

Phage ϕA1122 is a virulent phage of *Y. pestis* which at 37 °C is also active on *Y. pseudotuberculosis* (15). The ϕA1122 genome consists of 37,555 bp including direct terminal repeats of 148 bp. Fifty one gene products have been predicted.

T3/T7-Like Phages

ϕA1122 reveals a strong relationship to the sequenced yersiniophage, ϕYe03-12 (see below), and to coliphages T7 and T3 (see Chapter 20). Aside from some genes that have no positional counterpart in these other phages, the ϕA1122 genome is 89% and 73% identical to the genomes of T7 and T3, respectively. Moreover, the strong relationship is corroborated by similar or identical promoters, transcriptional terminators, terminal repeats, and the origins of replication. At the protein level, most ϕA1122 proteins exhibit more than 80% identity with the homologous T7 proteins. Three and four ϕA1122 proteins are even identical to their T7 and T3 counterparts, respectively. Interestingly, almost one quarter of the ϕA1122 genome (genes 15 to 19), coding for about half of the morphogenic functions, is 99.8% identical to T3. The G-C content of the corresponding T3 region is significantly less than that of the remainder of the T3 genome and much closer to that of the ϕA1122 genome. As the first 26 kb of the T3 genome, up to gene 15, is approximately 90% identical to the *Y. enterocolitica* phage ϕYeO3-12, it has been suggested that phage T3 might have been arisen by recombination of two yersiniophages (15).

Host Range Divergence

Although phage T3 does not infect *Yersinia*, the gp17 tail fiber proteins of T3 and ϕA1122, determining the host range of these phages, are nearly identical. It could be demonstrated that besides *Yersinia*, phage ϕA1122 is able to plate on *Escherichia coli* K12, albeit with a much lower efficiency. The phages released from *E. coli* showed an expanded host range infecting at the same efficiency both *Y. pestis* and *E. coli*. Sequence analysis revealed a single mutation in gene *17* of phage ϕA1122. This mutation leads to an amino acid exchange at position 523 near the C-terminus of the protein, which interacts with the bacterial cell surface. The same amino acid is present at the corresponding position in the tail fiber protein of T3. These data indicate that despite the fact that T3 is a specific phage of *E. coli*, while ϕA1122 mainly infects *Y. pestis*, it seems very likely that one or even two parents of T3 are yersiniophages.

Yersiniophage ϕYeO3-12

Isolation of the Phage

The *Y. enterocolitica* serotype O:3 (YeO3)-specific bacteriophage, ϕYeO3-12, was isolated in 1988 from the raw incoming sewage of the Turku City sewage treatment plant (54). A phage of identical characteristics was isolated from the sewage in 2000, indicating that the phage is stably colonizing the sewage system (M. Skurnik, unpublished observations).

Phage Characteristics

In electron micrographs, the ϕYeO3-12 particles have approximate dimensions of 57 nm for the head and 15 × 8 nm for the tail. Some capsids show pentagonal outlines, indicating their icosahedral nature. This morphology classifies ϕYeO3-12 to the family *Podoviridae* (35) and in Bradley's classification to type C (6) (see chapter 2 for a review of phage classification). When growing on *Y. enterocolitica* strain YeO3, ϕYeO3-12 has eclipse and latent periods of 15 and 25 min, respectively, followed by a short rise period of 10 min. The resulting lysis of the host cell produces a burst size of 100–140 phage progeny (41) (for an overview of phage infection-timing and burst-size characters see chapter 5).

Phage Receptor

The lipopolysaccharide (LPS) O-antigen (a homopolymer of 6-deoxy-L-altrose) of YeO3 is the phage receptor since (i) the phage-resistant YeO3 strains lack the O-antigen and (ii) the phage is able to infect *E. coli* strains expressing the cloned O-antigen of YeO3 (1). In addition to serotype O3, phage ϕYeO3-12 infects other *Y. enterocolitica* serotypes (such as O:1 and O:2) carrying 6-deoxy-L-altrose in their O-antigen. Also, *Y. frederiksenii* serotype O:3 and *Y. mollaretii* serotype O:3 were found to be phage-sensitive (41).

Nucleotide Sequence of ϕYeO3-12 (42)

The linear DNA of ϕYeO3-12 is 39,600 bp in size (GenBank/EMBL/DDBJ database accession No. AJ251805). The genome has an overall G-C content of 50.6%, which is close to that of its host, 48.5±1.5% (7). Altogether 58 genes were identified from the sequence (figure 34-1) (42). The nucleotide sequence shows striking similarity to that of phage T3 with an overall sequence identity of 84%. It is also approximately 70% identical to that of phage T7, and therefore is also closely related to the *Y. pestis* phage ϕA1122 (discussed above). Most of the 58 genes have counterparts in the genomes of phages ϕA1122, T3, and T7, and the genes are organized in the same order. The nucleotide sequences of phages ϕYeO3-12, T3, and T7 were compared and the results plotted using the DOTPLOT program of GCG (figure 34-2), showing that the three genomes align over their entire lengths. In a few places the identity lines are broken by upward or downward shifts caused by the presence or absence of a gene in the compared genomes. The high similarity between the ϕYeO3-12 and T3 genomes shows that they belong to the same lineage of phages that have developed different host specificities.

Promoters

The ϕYeO3-12-specific promoter is identical to that of phage T3 and consists of a 23 bp sequence (AATTAACCCTCACTAAAGGGAGA) that runs from −17 to +6 relative to the transcription start. ϕYeO3-12 has a total of 15 promoter motifs located in the genome in the same relative positions as the T3 promoters (figure 34-1).

Terminal Repeats

The genome contains identical direct terminal repeats of 232 bp that are 87% identical to the 230 bp terminal repeats of T3 and 56% identical to the 160 bp repeats of T7. The sequences at the beginning and end of the terminal repeats of the three phages are more similar than the sequences in the middle, indicating that the mechanisms of maturation of the DNA ends might be similar.

The ϕYeO3-12 Genes

Similar to the T3 and T7 genomes, ϕYeO3-12 genes can be divided into three classes: I (early), II (middle), and III

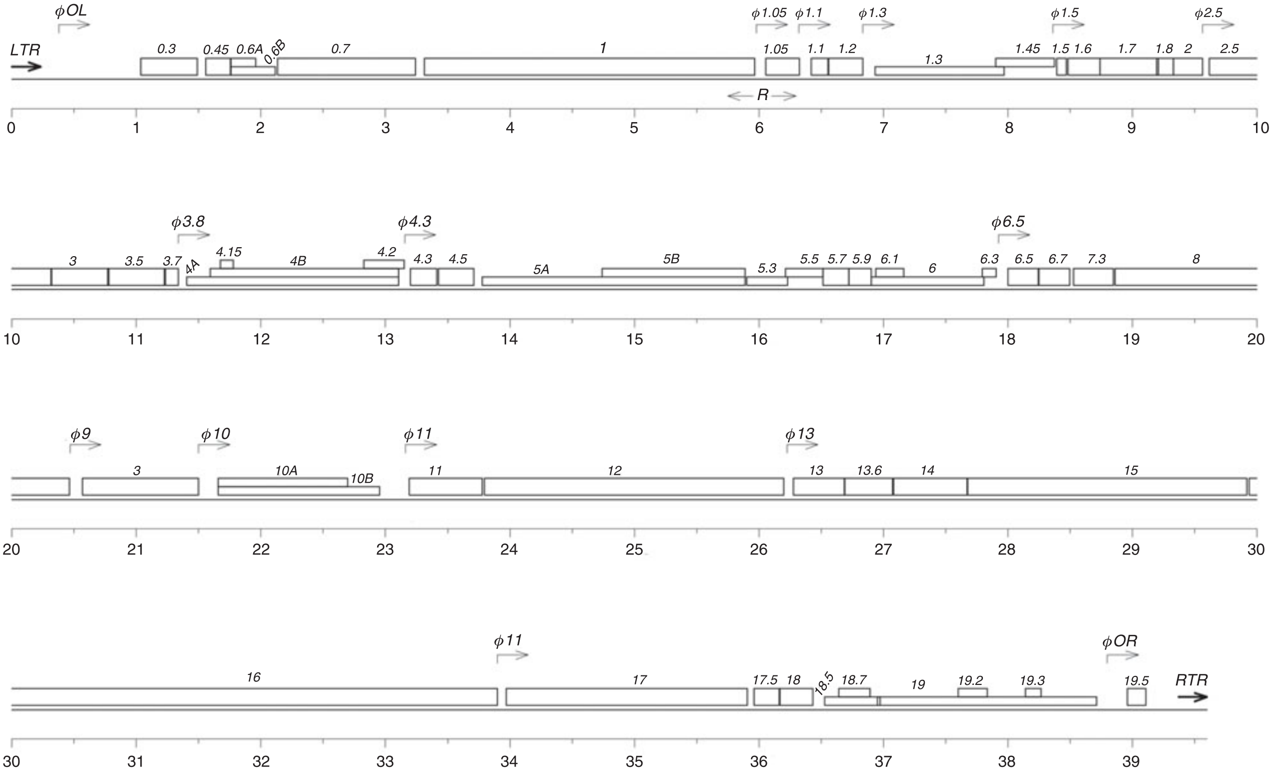

ϕOL
LTR
0.3
0.45
0.6A
0.6B
0.7
1
ϕ1.05
ϕ1.1
ϕ1.3
1.05
1.1
1.2
1.3
R
ϕ1.5
1.45
1.5
1.6
1.7
1.8
2
ϕ2.5
2.5
0
1
2
3
4
5
6
7
8
9
10
3
3.5
3.7
4A
ϕ3.8
4.15
4B
4.2
ϕ4.3
4.3
4.5
5A
5B
5.3
5.5
5.7
5.9
6.1
6
ϕ6.5
6.3
6.5
6.7
7.3
8
10
11
12
13
14
15
16
17
18
19
20
ϕ9
3
ϕ10
10A
10B
ϕ11
11
12
ϕ13
13
13.6
14
15
20
21
22
23
24
25
26
27
28
29
30
16
ϕ11
17
17.5
18
18.5
18.7
19
19.2
19.3
ϕOR
19.5
RTR
30
31
32
33
34
35
36
37
38
39

Figure 34-1 Genetic organization of the genome of bacteriophage ϕYeO3-12. The genes are represented by open boxes and the names of the genes are given above. Overlapping genes are drawn on top of each other. The locations of the promoters are indicated by bent arrows. LTR and RTR are the left and right terminal repeats, respectively. The primary origin of replication is indicated by R and arrows. T_E and Tø are transcription termination sites of host and phage ϕYeO3-12 RNA polymerases, respectively. The scale shown below the genes is in kilobases.

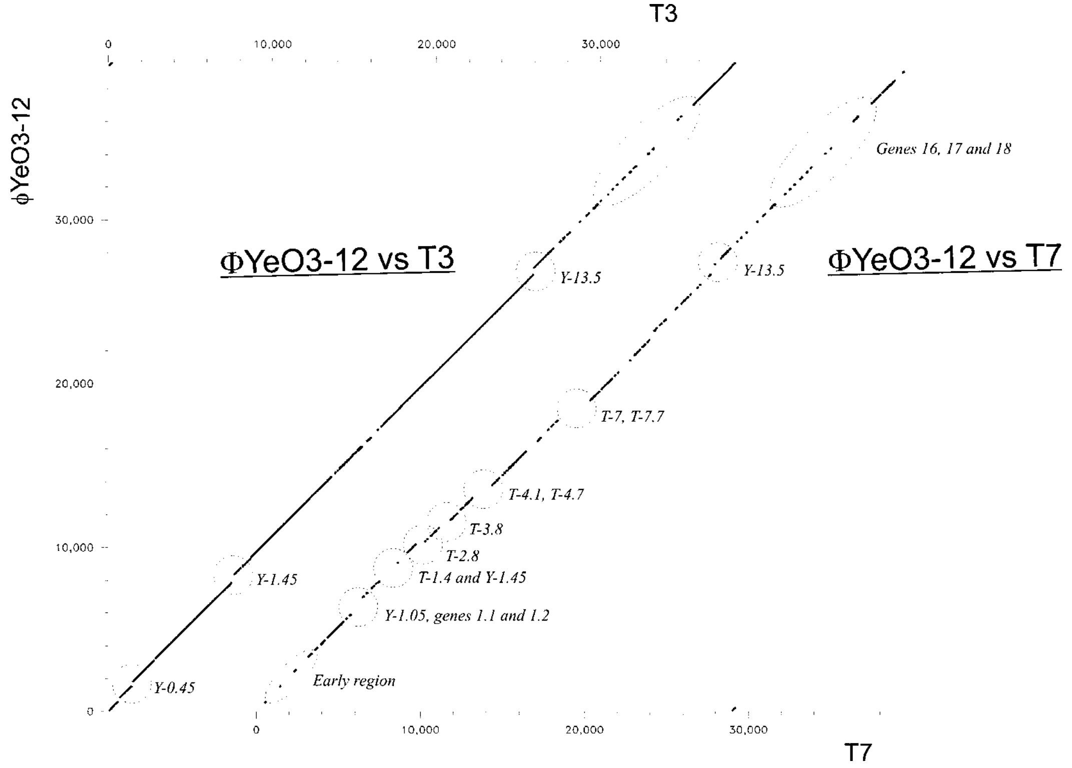


Figure 34-2 Pairwise genomic comparisons. Comparison of phage *ϕ*YeO3-12 with the phage T3 (on the left) and phage T7 (on the right). Pairwise comparisons were made using the COMPARE and DOTPLOT programs. Sequences were scanned using a 50 nucleotide window and each data point in the plot represents 80% identity or more between the sequences. Regions with gross sequence differences between genomes are circled with dotted lines and the gene insertions are indicated (*Y-nn* and *T-nn* indicate the presence of *ϕ*YeO3-12 phage-specific and T7 phage-specific genes, respectively).

(late). For most of the ϕYeO3-12 genes the functions are predicted based on similarity to T3 and T7 genes. In a random transposon mutagenesis approach, insertions into genes *0.45, 0.7, 1.1, 1.3, 1.6, 3.5, 3.7, 4.5,* and *5.5* were tolerated, indicating that their function is not absolutely essential for the viability of the phage (31). No insertions into the late genes were recovered.

Early Genes (Genes 0.3–1.45)

The gene products (gp) of these genes are involved in controlling the functions of the host and redirecting its metabolism to benefit phage propagation. Functions of the gene *0.45, 0.6A, 0.6B, 1.05,* and *1.1* products are not known. Gp0.3 is S-adenosyl-L-methionine hydrolase, which degrades the methyl group donor in the host (56). In agreement with this, no modified nucleotides were detected in ϕYeO3-12 DNA by high-performance liquid chromatography analysis (unpublished). Gp0.7 is a protein kinase; gp1, the RNA polymerase; gp1.2, a deoxyguanosine triphosphohydrolase inhibitor; and gp1.3, DNA ligase. Gp1.45 shows similarity to HNH endonucleases.

Middle Genes (Genes 1.5–6.3)

The middle gene products are involved in phage DNA replication. Functions of the gene *1.5, 1.6, 1.7, 1.8, 4.15, 4.2, 4.3, 4.5, 5.3, 5.5, 5.7, 5.9, 6.1,* and *6.3* products are not known. Gp2 is an inhibitor of the host RNA polymerase; gp2.5, a single-stranded DNA binding protein; gp3, an endonuclease; gp3.5, a lysozyme; gp3.7, a protein similar to a number of trypsin inhibitors (3); gp4A, a DNA primase-helicase; gp4B, a primase; gp5A, a DNA polymerase; and gp6, an exonuclease.

Late Genes (Genes 6.5–19.5)

The late gene products are either structural components of the phage particle or are involved in packaging the genome. Functions of the gene *6.5, 19.2, 19.3,* and *19.5* products are not known. Gp6.7 and gp7.3 are involved in adsorption of phage particles to the host and determining the host range, respectively. Gp8 is a head–tail connector; gp9, scaffolding protein; gp10A, major capsid protein; and gp10B, minor capsid protein. Gp11 and gp12 are tubular tail proteins, A and B, respectively. Gp13, gp14, gp15, and gp16 are internal phage head proteins, and gp17, the tail fiber protein. Gp13.5, which is specific for phage ϕYeO3-12, may be an endonuclease. Gp17.5 is a holin (holins are reviewed in chapter 10). Gp18 and gp19 are the small and large subunits of the DNA packaging protein, respectively. Gp18.5 and 18.7 are involved in host cell lysis.

Host Recognition

The phage tail fibers are required to confer infectivity of the phage particles. T3 tail fiber is a very stable trimer of gp17 and the trimer formation proceeds independently from its attachment to the tail (24, 25). Orientation of gp17 in the tail fiber is such that the N-terminus is proximal to the phage particle with subunits oriented coaxially to each other (26, 57). The counterpart(s) to which the N-terminus of gp17 binds is not known. The C-terminus of gp17 recognizes the receptor on the bacterial surface (57). Both T3 and T7 tail fibers use the outer core region of the *E. coli* lipopolysaccharide to initiate adsorption. Gene *17* of ϕYeO3-12 encodes a protein of 645 amino acid residues, more than 100 residues larger than its T3 and T7 homologs. The N-terminal 150 amino acids of ϕYeO3-12 gp17 show marked (78.7%) similarity to the N-terminal part of gp17 of both T3 and T7, though the C-terminal two thirds of the tail fiber proteins are much more different (ca. 31% similarity). The differences between the C-terminal regions of ϕYeO3-12 gp17 and T3 or T7 gp17 may thus be a reflection of host specificity.

A recombinant T3 phage that carries the gene *17* of ϕYeO3-12 is able to infect YeO3-c (43). The T3 recombinants were not stable, indicating that the ϕYeO3-12 gp17 does not attach firmly to the T3 particle and that additional mutations would be required to stabilize the recombinants. Nevertheless, the gene swapping proves that gp17 plays a major role in determining the host range of T7-group phages. However, other gene products, such as the major capsid protein gp10, may also play a role. Monospecific anti-gp10 antibodies are able to neutralize ϕYeO3-12 and also T3, although with lower efficiency (41), indicating that gp10 may have a role in the initiation of the infection.

Yersiniophage PY54

Isolation of the Phage

PY54 is a temperate phage isolated from a nonpathogenic *Y. enterocolitica* O:5 strain by induction with mitomycin C (45). As well as O:5 strains, pathogenic strains belonging to the serogroup O:5,27 are also infected. Plaques are formed at 28 °C but not at 37 °C.

Phage Characteristics

The PY54 virion has a phage λ-like morphology and the phage was classified in the *Siphoviridae* family (see chapter 2). Apart from an obviously identical phage, PY54 showed no significant DNA homology to other temperate yersiniophages (45). The phage genome is a sticky-ended double-stranded DNA of 46 kb with 10 nucleotide (′-GGG ACA GGC A-3′ and 5′-TGC CTG TCC C-3′) overhangs at the 3′-ends (19). 3′-protruding ends are quite uncommon for phages of Gram-negative bacteria, and have been previously reported for only a few other phages, such as *Pseudomonas aeruginosa* phage D3 (53), *Burkholderia mallei* phage ϕE125 (60), and the *E. coli* phages HK022, HK97, and ϕP27 (22, 51).

Another unusual property of PY54 is that its prophage is not integrated into the bacterial chromosome but replicates as a linear low-copy-number plasmid. The comparison of the plasmid prophage with phage DNA isolated from particles revealed that these molecules have the same size and are nearly 50% circularly permuted. The plasmid ends are covalently closed hairpin structures similar to the telomeres found in *E. coli* phage N15 (52) (phage N15 is reviewed in chapter 28). Until now N15 was unique in terms of its replication as a linear plasmid prophage with terminal hairpins. Similar to PY54, the N15 genome is a circular permutation of its prophage. In the center of the genome, N15 contains the *telN* gene, encoding a protelomerase, which cleaves a 56 bp palindrome (*telRL*) located close to *telN* (14). The staggered cuts yield overhanging single strands, which are self-complementary and are able to snap back. The DNA strands are then rejoined by the protelomerase. Through this cleaving/joining activity, the N15 plasmid telomeres are generated, each of which is composed of one DNA strand of the palindrome. In the central part on the PY54 genome, a 42 bp palindrome and an adjacent open reading frame (*tel*) were identified that showed homologies to the aforementioned sequences of N15 (figure 34-3). The PY54 *tel* gene was cloned and expressed in *E. coli* and a 77 kDa protein was obtained and purified (19). The protein was demonstrated to cleave the 42 bp palindromic sequence of PY54, indicating that the phage uses the same principle for the generation of the plasmid ends as N15. However, in vivo studies with PY54 showed that adjacent DNA sequences from the phage are also important for successful cleavage. These sequences comprise another 15 bp inverted repeat flanking the palindrome that is important for cleavage in infected bacteria (figure 34-3). The data strongly suggest an essential role of the protelomerases for conversion of the phage genomes into the linear plasmid prophages. Moreover, as already demonstrated for the N15 protelomerase, this enzyme is also essential for the replication and maintenance of linear N15 plasmids (48).

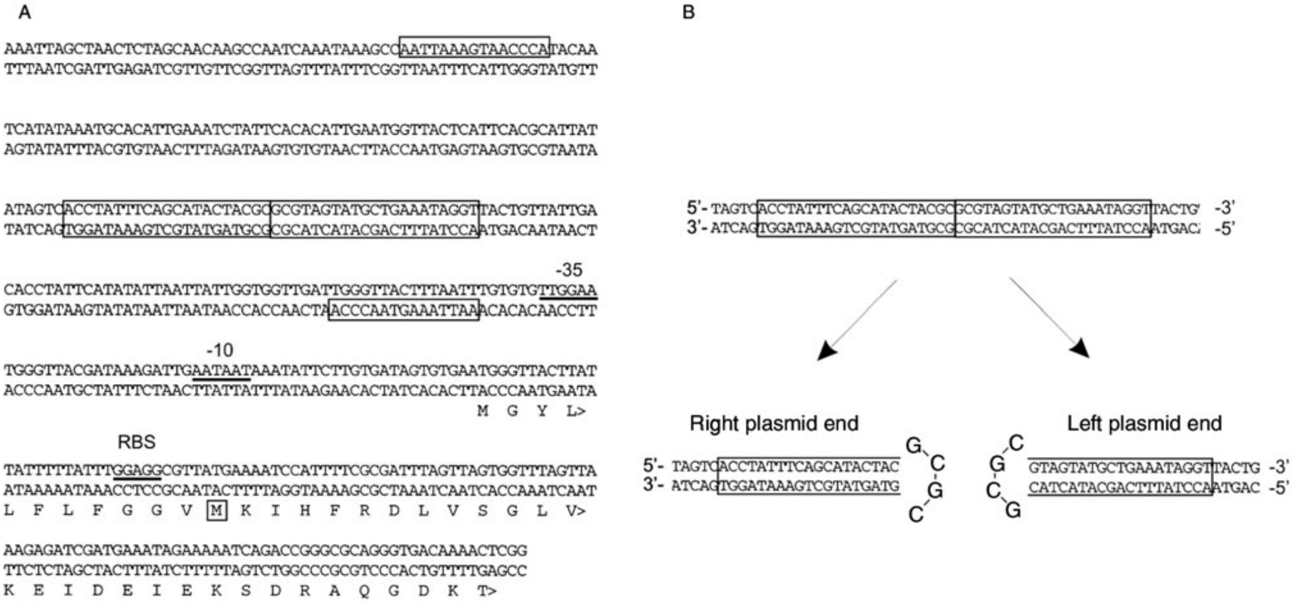

A
AAATTAGCTAACTCTAGCAACAAGCCAATCAAATAAAGCCAATTAAAGTAACCCATACAA
TTTAATCGATTGAGATCGTTGTTCGGTTAGTTTATTTCGGTTAATTTCATTGGGTATGTT
TCATATAAATGCACATTGAAATCTATTCACACATTGAATGGTTACTCATTCACGCATTAT
AGTATATTTACGTGTAACTTTAGATAAGTGTGTAACTTACCAATGAGTAAGTGCGTAATA
ATAGTCACCTATTTCAGCATACTACGCGCGTAGTATGCTGAAATAGGTTACTGTTATTGA
TATCAGTGGATAAAGTCGTATGATGCGCGCATCATACGACTTTATCCAATGACAATAACT
-35
CACCTATTCATATATTAATTATTGGTGGTTGATTGGGTTACTTTAATTTGTGTGTTGGAA
GTGGATAAGTATATAATTAATAACCACCAACTAACCCAATGAAATTAAACACACAACCTT
-10
TGGGTTACGATAAAGATTGAATAATAAATATTCTTGTGATAGTGTGAATGGGTTACTTAT
ACCCAATGCTATTTCTAACTTATTATTTATAAGAACACTATCACACTTACCCAATGAATA
M G Y L>
RBS
TATTTTTATTTGGAGGCGTTATGAAAATCCATTTTCGCGATTTAGTTAGTGGTTTAGTTA
ATAAAAATAAACCTCCGCAATACTTTTAGGTAAAAGCGCTAAATCAATCACCAAATCAAT
L F L F G G V M K I H F R D L V S G L V>
AAGAGATCGATGAAATAGAAAAATCAGACCGGGCGCAGGGTGACAAAACTCGG
TTCTCTAGCTACTTTATCTTTTTAGTCTGGCCCGCGTCCCACTGTTTTGAGCC
K E I D E I E K S D R A Q G D K T>
B
5'- TAGTCACCTATTTCAGCATACTACGCGCGTAGTATGCTGAAATAGGTTACTG -3'
3'- ATCAGTGGATAAAGTCGTATGATGCGCGCATCATACGACTTTATCCAATGAC -5'
Right plasmid end
5'- TAGTCACCTATTTCAGCATACTAC
3'- ATCAGTGGATAAAGTCGTATGATG
G C G C
Left plasmid end
C G C G
GTAGTATGCTGAAATAGGTTACTG -3'
CATCATACGACTTTATCCAATGAC -5'

Figure 34-3 Structures of the PY54 plasmid ends. A: Nucleotide sequence of the DNA region upstream from the *tel* gene (AJ348844). Boxed are the 42 bp palindrome, its flanking 15 bp inverted repeat and the putative start codon, ribosome binding site (RBS), and putative promotor sequences of the *tel* gene. B: DNA sequences of the right and left plasmid ends.

Guanine + Cytosine (G-C) Content of PY54

The complete nucleotide sequence of phage PY54 has been determined (20). The genome has a size of 46,339 bp with an average G-Content of 44.6%, which is slightly lower than the G-C content reported for the *Y. enterocolitica* genome. However, the G-C content of different PY54 open reading frames (ORFs) vary, ranging from 27.6% to 58%. With regard to the G-C values, the phage genome can be divided into 3 regions. The ORFs 1–18 (41.4–48.8%) are separated from ORFs 20–32 (39.4–58%) by ORF 19 (27.6%). Similarly the latter ORFs are separated from ORFs 34–67 (34.5–52.3%) by ORF 33 (31.2%). Is it possible that the ORFs 19 and 33, with a much lower G-C content, are "morons" (17) (see chapter 27 for further discussion of phage morons).

Sequence Analysis of the PY54 Genome

Sixty-seven candidate genes with good coding potential and a start codon (63 with AUG, two with TTG, two with GTG) were assigned (table 34.1). For most of the ORFs, a plausible Shine–Dalgarno sequence was identified. Fifty-five putative PY54 genes are transcribed rightward (on the genetic map) and 12 genes leftward. Bioinformatic analysis revealed 44 gene products for which homologs were found. The *tel* site of PY54 is located on the genome similarly to the *tel* site in N15 and the attachment site, *attP*, in phage λ. In analogy to these phages, the PY54 genome was divided into a left and a right arm, separated by the *tel* site (figure 34-4).

The left PY54 Arm

As is generally the case for lambdoid phages (see chapter 27), the left arm of phage PY54 apparently contains "late genes," coding for structural and assembly proteins. Again consistent with lambdoid phages, the right arm consists mainly of regulatory genes important for plasmid replication, genes required for host cell lysis, and genes for control of DNA methylation. Nevertheless, in contrast to phage N15 whose left arm is closely related to λ-like phages, the actual left-arm genes of phage PY54 are much less similar to this family.

The products of ORF 1 and ORF 2, of the left arm are probably the small and large subunits of the phage terminase, respectively. While the predicted ORF 1 product shows only weak homology to the terminase small subunits of the *B. mallei* phage ϕE125 and *E. coli* phage VT2-Sakai, the ORF 2 product reveals significant similarity to the terminase large subunits of the phages ϕP27, ϕE125, and D3. As noted above, these phages of Gram-negative bacteria share the property of 3′ protruding DNA ends. The ORFs 3 to 27 of PY54 are suspected to code for proteins required for virion head and tail assembly but up to now only eight of the deduced gene products could be assigned by their similarities to known structural proteins of other phages or by functional analysis. The PY54 major capsid protein is most probably encoded by ORF 5. Its deduced product is very similar to the major capsid protein of phage ϕE125. Moreover, the analysis of the PY54 structural protein profile revealed one major protein band of about 38 kDa (45).

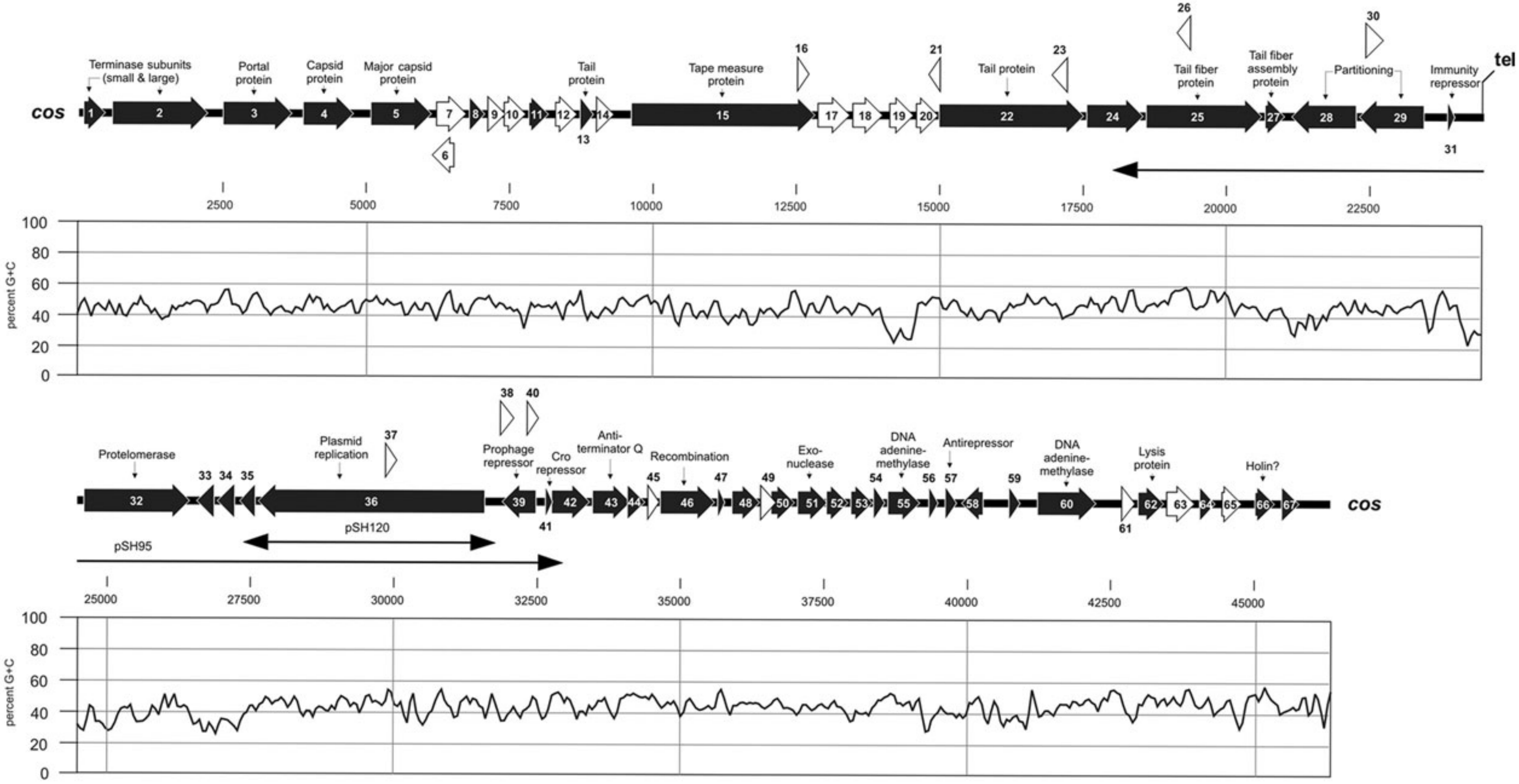

Terminase subunits (small & large)
Portal protein
Capsid protein
Major capsid protein
Tail protein
Tape measure protein
16
21
Tail protein
23
26
Tail fiber protein
Tail fiber assembly protein
30
Partitioning
Immunity repressor
tel
cos
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
17
18
19
20
22
24
25
27
28
29
31
2500
5000
7500
10000
12500
15000
17500
20000
22500
percent G+C
100
80
60
40
20
0
Protelomerase
Plasmid replication
37
38
40
Prophage repressor
Cro repressor
Anti-terminator Q
Recombination
Exo-nuclease
DNA adenine-methylase
Antirepressor
DNA adenine-methylase
Lysis protein
Holin?
32
33
34
35
36
39
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
cos
pSH95
pSH120
25000
27500
30000
32500
35000
37500
40000
42500
45000
percent G+C
100
80
60
40
20
0

Figure 34-4 Arrangement of open reading frames (ORFs) on the PY54 genome. The upper part shows the left arm, separated from the right arm (lower part) by the *tel* site. The ORFs shown in black indicate similarity to known genes as follows: ORF 31, 32, 34–36, 39, 41, 43, 44, 52, 54, 57, 59, 64 (N15), ORF 13, 22, 25, 33, 50 (lambdoid phages), ORF 1–5, 8, 11, 27, 48, 51, 53, 55, 60, 66 (phages of other Gram-negative bacteria), ORF 15 and 67 (phages of Gram-positive bacteria), ORF 24, 28, 29, 58 (*Yersinia*), ORF 42, 46, 47, 56, 62 (other bacteria): see table 34-1 for details. ORFs with no database matches are shown in white. The arrows indicate the positions of the smallest linear (pSH95) and circular (pSH120) miniplasmid. Scans of G-C content across the phage genome are shown below the gene map. See thebacteriophages.org/frames_0340.htm for a color version of this figure.

This protein was analyzed by peptide mass fingerprinting, demonstrating that it is indeed encoded by ORF 5 (20).

There are two other putative PY54 products (from ORFs 3 and 4) that are closely related to phage ϕE125 structural proteins. The predicted ORF 3 product exhibits significant similarity to the portal protein and the ORF 4 product to the capsid assembly protein of ϕE125. Therefore it is conceivable that PY54 has acquired a gene cluster by recombination with a ϕE125-related phage that is comprised of ORF 3 (or ORF 2) through ORF 5. Two of the predicted PY54 proteins (22 and 25) are related to phage λ tail proteins (host specificity protein J and tail fiber protein, respectively). In addition, the putative ORF 13 product is similar to the major tail subunit of the λ-like phage HK97, but the predicted phage PY54 protein is much smaller than the related HK97 protein. The other assigned structural proteins (15 and 27) of PY54 show homologies to the tape measure protein of the *Lactococcus lactis* phage TP901-1 and the tail fiber assembly protein of phage APSE-1. A PY54 phage mutant harboring a kanamycin resistance gene within ORF 15 showed capsids containing DNA but no tails, indicating that this gene is essential for the assembly of the tail.

There are 17 ORFs in the PY54 left arm which might encode proteins involved in phage assembly but for which no function could be determined. This is in marked constrast to the genome of phage N15, in which genes 2 to 21 are correlated with phage λ genes with up to 90% identity between the predicted amino acid sequences of the head gene products (50). Similar to N15, the products of two genes (ORF 28 and ORF 29) found at the right end of the left arm of phage PY54 are homologs of proteins responsible for plasmid partition. The strongest homologies were found to the SpyA and SpyB proteins encoded by the *Y. enterocolitica* virulence plasmid, pYV, and, to a slightly lesser degree, to the SopA and SopB proteins of phage N15, which in turn are closely related to SopA/B protein of the F plasmid. These proteins interact with centromere-like sites (*spyC* or *sopC*) comprising direct or inverted repeats. In the F plasmid, *sopC* is located downstream from *sopA/B* and contains multiple copies of the probable recognition site (5′-TGGGACCnnGGTCCCA; n denotes variable bases) (32). Scattered on a 13 kb fragment of N15, four inverted repeats resembling the sequence shown above were identified and demonstrated to act as centromeres (16, 46). On plasmid pYV, *spyC* is located downstream from *spyA/B* but this sequence is composed of direct repeats harboring a related recognition site (21, 55). The analysis of the phage PY54 DNA revealed that this phage contains eight identical copies of the *sopC* recognition site of the F plasmid. The inverted repeats are scattered over the whole PY54 genome. Although a function of these repeats as centromeres has not been demonstrated yet, it appears that the centromere-like sites of PY54 are more similar to *sopC* of the F plasmid and phage N15 than to *spyC* of the *Y. enterocolitica* virulence plasmid, pYV.

The PY54 Right Arm

The right arm of the PY54 genome is considerably more similar to N15 DNA than the left arm. Of 36 ORFs identified on the right arm of phage PY54, 13 deduced gene products (of genes *32*, *34*, *35*, *36*, *39*, *41*, *43*, *44*, *52*, *54*, *57*, *59*, and *64*) show the closest relationship to phage N15 proteins (table 34-1). In addition, two other putative PY54 products (of genes *48* and *51*) are also similar to N15 proteins. Most of the related N15 proteins are important for plasmid replication, are part of the immunity or anti-immunity system of N15, or their function is unknown. As noted above, the protelomerase of PY54 is closely related (overall similarity 60%) to the protelomerase, TelN, of phage N15. Furthermore, both proteins contain a sequence motif at the C-terminus which is also present in several integrases and proposed to belong to the active center of these enzymes (14). It is conceivable that the PY54 protelomerase, like its N15 counterpart, is involved in plasmid replication by processing of replicative intermediates.

Another probable replication protein of phage PY54 is specified by ORF 36. This is the largest gene on the PY54 genome and its predicted product shows strong homology to and is nearly the same size as RepA of N15. RepA is suggested to be a multifunctional replication protein comprising primase, helicase, and origin binding activities. This was concluded by similarities to primases of conjugative plasmids and the primase of phage P4 (44, 61) (phage P4 is reviewed in chapter 26). An essential role of RepA in plasmid replication was also demonstrated by the isolation of N15 and PY54 miniplasmids. While a circular 4.5 kb miniplasmid of PY54 composed of ORF 36 and part of ORF 35 retains replicative competence (19), a 4.2 kb DNA fragment of N15 containing the *repA* gene and 10 bp of *cB* is sufficient to drive the replication of a circular N15 plasmid (49). These observations suggest that the *rep* genes of PY54 and N15 have the same function and that the plasmid *ori* sites might be located within the *rep* genes.

The sequence analysis of the PY54 genome revealed a number of ORFs (genes *39*, *41*, *43*, *44*, *47*) whose predicted products are related to N15 proteins encoded by the primary immunity region, *immB* (34). ORF 39 of PY54 apparently represents the prophage repressor. The predicted protein is similar to protein CB of phage N15, which in turn is related to lambdoid repressors. By cloning ORF 39 in *Yersinia*, it has been shown that this gene confers resistance against lytic infection (20). The ORFs 41 and 43 are located upstream from ORF 39. Their products are similar to the putative N15 antirepressor, Cro, and the transcription antiterminator, Q, respectively.

In contrast to phage N15, the genes *41* and *43* of phage PY54 are not arranged in an operon but instead are separated by about 1 kb. The probable function of ORF 41 as an antirepressor has been previously reported. After cloning this gene in *Yersinia*, a 10^2- to 10^3-fold increased number

Table 34-1 Bacteriophage PY54 ORF Analysis

ORF	Start	Stop	Strand	Mass (kDa)	PI	Functional assignment (Related sequence)	% Identity
1	61	504	+	16.1	8.8	Terminase small subunit (ϕE125)	23
2	568	2301	+	64.3	6.8	Terminase large subunit (ϕP27, SfV, ϕE125, D3)	55/53/51/40
3	2492	3751	+	47	8.5	Portal protein (ϕE125, ϕP27, HK022)	49/35/29
4	3906	4817	+	33	4.9	Capsid assembly protein (ϕE125)	43
5	5068	6168	+	38.7	5.1	Major capsid protein (ϕE125)	48
6	6574	6158	–	14.1	4.1		
7	6234	6800	+	19.7	4.4		
8	6778	7110	+	12.5	4.1	Unknown (SfV ORF 6 product)	30
9	7110	7463	+	13.1	9.2		
10	7402	7824	+	16.2	5.5		
11	7830	8237	+	14.9	9.8	Unknown (D3 ORF 14 product, HK97 gp10)	29/27
12	8280	8732	+	15.8	4.4		
13	8708	8992	+	9.6	4.7	Major tail subunit (HK97)	35
14	9001	9372	+	13.3	6.3		
15	9605	12868	+	115.8	6.5	Tape measure protein (TP901-1)	26
16	12522	12749	+	8.9	9.3		
17	12872	13468	+	22.1	9.3		
18	13468	14052	+	21.2	5.8		
19	14113	14583	+	18	6.4		
20	14605	15006	+	14.5	7.2		
21	15036	14794	–	8.7	10.1		
22	14999	17542	+	92.9	4.7	Tail protein (λ host specificity protein)	27
23	17269	16949	–	12	9.6		
24	17553	18575	+	36.4	8.8	Unknown (*Y. pestis* hypothetical phage protein)	32
25	18588	20660	+	71.1	4.6	Tail fiber protein (λ)	44
26	19402	19139	–	8.5	3.6		
27	20660	21067	+	15.1	4.0	Tail fiber assembly protein (APSE-1)	36
28	22316	21135	–	43.9	8.0	Partitioning (*Y. ent.* SpyB, N15 SopB)	56/54
29	23482	22313	–	43.4	5.3	Partitioning (*Y. ent.* SpyA, N15 SopA)	85/70
30	22418	22765	+	13.2	8.4		
31	23858	24022	+	6.1	11.1	Inhibitor of cell division (N15 Icd)	59
32	24604	26490	+	71.8	5.6	Protelomerase (N15)	40
33	26915	26544	–	14.3	4.6	Unknown (P22 Eaf protein)	55
34	27257	26928	–	12.9	8.8	Unknown (N15 gp33)	45
35	27615	27304	–	12.2	5.1	Unknown (N15 gp35)	35
36	31623	27631	–	148.8	7.8	DNA replication (N15 RepA)	49
37	29846	30091	+	9.1	12.2		
38	31862	32131	+	10.6	8.4		
39	32524	31877	–	24.1	8.2	Prophage repressor (N15 CB)	35
40	32319	32564	+	8.9	11.7		
41	32629	32832	+	7.9	10.1	Cro repressor (N15)	31
42	32747	33451	+	26.2	6.7	Unknown (*X. fastidiosa* hypothetical protein)	29
43	33452	34165	+	26.6	9.4	Antiterminator Q (N15)	31
44	34077	34427	+	13.3	9.1	Unknown (N 15 QD1)	52
45	34430	34672	+	9.0	10.1		
46	34654	35631	+	36.2	4.9	Unknown (*P. multocida* recombination associated protein)	26
47	35634	35834	+	7.6	9.9	Unknown (N15 QD1)	35
48	35883	36410	+	19.5	4.4	Unknown (CP-933O gene Z2097 product)	43
49	36407	36706	+	10.9	3.9		
50	36571	37038	+	17.6	10.3	Unknown (HK022 hypothetical protein)	50
51	37042	37614	+	20.7	4.8	Exonuclease (CP-933O, CP-933P)	53/52
52	37538	37978	+	16.6	5.5	Unknown (N15 gp41)	40
53	37972	38382	+	15.7	9.6	Unknown (CP-933V gene Z3348 product)	46
54	38366	38596	+	8.4	10.5	Unknown (N15 gp47)	47
55	38596	39228	+	23.8	5.9	DNA adenine-methylase (SfV)	31
56	39315	39551	+	8.8	5.1	Unknown (*S. sonnei* ImpC)	42
57	39601	39864	+	10.1	5.5	Antirepressor (N15 AntA)	35
58	40325	39891	–	15.7	5.0	Unknown (*Y. pestis* hypothetical protein)	48
59	40722	40955	+	9.1	4.8	Unknown (N15 gp51)	28
60	41231	42304	+	40.9	6.1	DNA adenine-methylase (SfV, CP-933O, ϕP27)	60/60/58
61	42687	42953	+	9.9	6.5		
62	42953	43486	+	19.5	9.4	Lysin (*S. dysenteriae*, 933W)	67/64
63	43479	44000	+	18.6	8.2		
64	44029	44355	+	11.9	6.3	Unknown (N15 gp57)	60
65	44442	44837	+	14.3	4.9		
66	45025	45387	+	13.8	10.0	Unknown (SfV ORF 53 product)	59
67	45456	45830	+	13.9	7.0	Unknown (D29 gp 64)	51

of plaques were obtained upon infection with phage PY54 (20). Further upstream on the PY54 genome there are two other ORFs (44 and 47) with homologies to the *immB* region of N15. Their products share similarity to the ORF QD1 product of phage N15, whose function is unknown. Two of the probable PY54 gene products (31 and 57) show homologies to N15 proteins encoded by the anti-immunity locus *immA*, but the homology of the ORF 57 product is rather weak.

The *immA* locus of N15 is involved in the choice between the lytic and lysogenic cycle and contains three genes: *icd*, *antA*, and *antB* (47). The genes are part of an operon and are transcribed from right to left. Interestingly, ORF 31 and ORF 57 of PY54, which resemble *icd* and *antA*, respectively, are transcribed from left to right and are a long way apart. ORF 31 is located at the right end of the left PY54 arm adjacent to the plasmid partitioning genes, whereas ORF 57 is located on the right arm about 15 kb upstream from ORF 31. The N15 proteins Icd and AntA are an inhibitor of cell division and an antirepressor, respectively (47). An antirepressor function has also been shown for the *antA*-related PY54 ORF 57 (20). For that reason, PY54 harbors at least two genes (ORF 41 and ORF 57) that might act as antirepressors.

The functions of the remaining putative PY54 proteins that are similar to N15 proteins cannot be predicted as yet. The products of the PY54 ORFs 34, 35, 52, 54, 59, and 64 make only convincing database matches to the N15 products 33, 35, 41, 47, 51, and 57, respectively, whose functions are still unknown. However, the fact that the order and orientation of the corresponding genes in PY54 and N15 are identical and that, with the exception of the ORF 54 product, the other PY54 products are similar in size to their N15 counterparts suggests a strong relationship.

Besides ORFs encoding N15-related proteins, the right PY54 arm contains a number of ORFs whose deduced products resemble proteins of other organisms. There are two ORFs (55 and 60) which obviously specify adenine-specific methyltransferases. The strongest similarities were detected to enzymes encoded by the *Shigella flexneri* phage SfV. ORF 51 may code for a product similar to exonucleases of cryptic *E. coli* prophages. The predicted product of ORF 46 yields database matches to recombination-associated proteins of *Pasteurella multocida* and *Haemophilus influenzae*. A probable lysis protein closely related to endolysins of *Shigella dysenteriae* and *E. coli* phage 933W may be encoded by ORF 62. Finally, the right arm of phage PY54 comprises some ORFs whose predicted products are similar to hypothetical or unknown proteins of other bacteria and phages (table 34.1).

Conclusion

The analysis of the PY54 genome reveals that this phage is not simply a derivative of phage N15 with an altered host range but a unique phage which has evolved independently. The genes on the left arm of PY54 are much more divergent from λ-like genes than those of N15. Although it can be assumed that this part of the PY54 genome comprises structural genes, many predicted gene products could not be assigned functions. In contrast, a significant number of genes located on the phage PY54 right arm show striking similarities to N15 genes. The probable products of these related genes are mainly involved in plasmid replication or have regulatory functions. Some of the genes have unknown functions. It can be suggested that the genes which are similar in PY54 and N15 were acquired by horizontal exchange. However, the right PY54 arm additionally contains a number of genes quite different from those of phage N15, confirming the genetic mosaicism demonstrated for lambdoid phages (22) (see chapter 27).

While this chapter was being completed, another linear plasmid prophage, KO2, was reported for *Klebsiella oxytoca* (13). Phage KO2 has not been propagated on indicator strains as yet. It has a genome of 51,601 bp, of which 64 genes have been predicted. The sequence analysis disclosed a strong relationship to N15. Some regions of the genome, for example genes *1* to *15* (head and tail shaft), *24* and *25* (partitioning), *26* (protelomerase), and *35* (replicase), are similar to genes of phage PY54. The data imply that KO2 has a similar life-style to N15 and PY54 and suggests that the number of phages replicating as linear plasmids with covalently closed ends will continue to increase in the next future.

References

1. Al-Hendy, A., P. Toivanen, and M. Skurnik. 1991. Expression cloning of the *Yersinia enterocolitica* O:3 *rfb* gene cluster in *Escherichia coli* K12. Microb. Pathog. 10:47–59.
2. Baker, P. M., and J. J. Farmer 3rd. 1982. New bacteriophage typing system for *Yersinia enterocolitica, Y. kristensenii, Y. frederiksenii*, and *Y. intermedia* correlation with serotyping, biotyping and antibiotic susceptibility. J. Clin. Microbiol. 15:491–502.
3. Beck, P. J., S. Gonzalez, C. L. Ward, and I. J. Molineux. 1989. Sequence of bacteriophage T3 DNA from gene 2.5 through gene 9. J. Mol. Biol. 210:687–701.
4. Bergan, T., and J. R. Norris. 1978. Bacteriophage typing of *Yersinia enterocolitica*. Methods Microbiol. 12:25–36.
5. Bottone, E. J. 1997. *Yersinia enterocolitica:* the charisma continues. Clin. Microbiol. Rev. 10:257–276.
6. Bradley, D. E. 1967. Ultrastructure of bacteriophage and bacteriocins. Bacteriol. Rev. 31:230–314.
7. Brenner, D. J., J. Ursing, H. Bercovier, A. G. Steigerwalt, G. R. Fanning, J. M. Alonso, and H. H. Mollaret. 1980. Deoxyribonucleic acid relatedness in *Yersinia enterocolitica* and *Yersinia enterocolitica*-like organisms. Curr. Microbiol. 4:195–200.

8. Calvo, C., J. Brault, J. M. Alonso, and H. H. Mollaret. 1981. New waterborne bacteriophages active on *Yersinia enterocolitica* . Appl. Environ. Microbiol. 4:235–38.
9. Calvo, C., and J. Brault. 1983. Lysogénie chez *Yersinia frederiksenii*, *Y. kristensenii* et *Y. intermedia*. Ann. Microbiol. Inst. Pasteur (Paris) 134A:183–188.
10. Calvo, C., R. Melis, J. Brault, J. M. Alonso, A. Ramos-Cormenzana, and H. H. Mollaret. 1986. Antagonism between *Yersinia intermedia* and *Yersinia enterocolitica* in water. Folia Microbiol. 31:167–173.
11. Calvo, C., J. Brault, M. A. Chalvignac, A. Ramos-Cormenzana, and H. H. Mollaret. 1986. Ability of *Yersinia enterocolitica* cells to inactivate bacteriophage I at 25 and 37 °C. Microbios 46:137–141.
12. Calvo, C., A. Fernandez, V. Bejar, and A. Ramos-Cormenzana. 1987. Differenciation rapide des sérovars O:3 et O:9 de *Yersinia enterocolitica* par certains bactériophages. Ann. Inst. Pasteur Microbiol. (Paris) 138:617–623.
13. Casjens, S. R., E. B. Gilcrease, W. Mun Huang, K. L. Bunny, M. L. Pedulla, M. E. Ford, J. M. Houtz, G. F. Hatfull, and R. W. Hendrix. 2004. The pKO2 linear plasmid prophage of *Klebsiella oxytoca*. J. Bacteriol. 186:1818–1832.
14. Deneke, J., G. Ziegelin, R. Lurz, and E. Lanka. 2000. The protelomerase of temperate *Escherichia coli* phage N15 has cleaving-joining activity. Proc. Natl. Acad. Sci. USA 97:7721–7726.
15. Garcia, E., J. M. Elliott, E. Ramanculov, P. S. G. Chain, M. C. Chu, and I. J. Molineux. 2003. The genome sequence of *Yersinia pestis* bacteriophage ΦA1122 reveals an intimate history with the coliphage T3 and T7 genomes. J. Bacteriol. 185:5248–5262.
16. Grigoriev, P. S., and M. B. Lobocka. 2001. Determinants of segregational stability of the linear plasmid-prophage N15 of *Eschericha coli*. Mol. Mircobiol. 42:355–368.
17. Hendrix, R. W., J. G. Lawerence, G. F. Hatfull, and S. Casjens. 2000. The origins and ongoing evolution of viruses. Trends Microbiol. 8:504–509.
18. Hertwig, S., A. Popp, B. Freytag, R. Lurz, and B. Appel. 1999. Generalized transduction of small *Yersinia enterocolitica* plasmids. Appl. Environ. Microbiol. 65:3862–3866.
19. Hertwig, S., I. Klein, R. Lurz, E. Lanka, and B. Appel. 2003. PY54, a linear plasmid prophage of *Yersinia enterocolitica* with covalently closed ends. Mol. Microbiol. 48:989–1003.
20. Hertwig, S., I. Klein, V. Schmidt, S. Beck, J. A. Hammerl, and B. Appel. 2003. Sequence analysis of the genome of the temperate *Yersinia enterocolitica* phage PY54. J. Mol. Biol. 331:605–622.
21. Iriarte, M., I. Lambermont, C. Kerbourch, and G. R. Cornelis. 1998. *Yersinia enterocolitica* plasmid pYVe227, complete sequence. GenBank accession no. AF102990.
22. Juhala, R. J., M. E. Ford, R. L. Duda, A. Youlton, G. F. Hatfull, and R. W. Hendrix. 2000. Genomic sequences of bacteriophages HK97 and HK22 pervasive genetic mosaicism in the lambdoid bacteriophages. J. Mol. Biol. 299:27–51.
23. Kasatiya, S., and H. W. Ackermann. 1986. Morphology of *Yersinia enterocolitica* phages. Ann. Inst. Pasteur Virol. 134E:59–69.
24. Kato, H., H. Fujisawa, and T. Minagawa. 1985. Genetic analysis of subunit assembly of the tail fiber of bacteriophage T3. Virology 146:12–21.
25. Kato, H., H. Fujisawa, and T. Minagawa. 1985. Purification and characterization of gene 17 product of bacteriophage T3. Virology 146:22–26.
26. Kato, H., H. Fujisawa, and T. Minagawa. 1986. Subunit arrangement of the tail fiber of bacteriophage T3. Virology 153:80–86.
27. Kawaoka, Y., K. Otsuki, and M. Tsubokura. 1982. Characteristics of *Yersinia enterocolitica* bacteriophages. Zentralbl. Bakt. Hyg. 253:102–109.
28. Kawaoka, Y., K. Otsuki, and M. Tsubokura. 1982. Temperature-dependent variation in the synthesis of the receptor for *Yersinia enterocolitica* bacteriophage XI. Zentralbl. Bakt. Hyg. 253:364–369.
29. Kawaoka, Y., K. Otsuki, and M. Tsubokura. 1983. Growth temperature-dependent variation in the bacteriophage-inactivating capacity and antigenicity of *Yersinia enterocolitica* lipopolysaccharide. J. Gen. Microbiol. 129:2739–2747.
30. Kawaoka, Y., T. Mitani, K. Otsuki, and M. Tsubokura. 1987. Isolation and use of eight phages for typing *Yersinia enterocolitica* O:3. J. Med. Microbiol. 23:349–352.
31. Kiljunen, S., H. Vilen, H. Savilahti, and M. Skurnik. 2003. Transposon mutagenesis of the phage ϕYeO3–12, pp. 245–248. *In* M. Skurnik, K. Granfors, and J. A. Bengoechea (eds.) The Genus *Yersinia*: Entering the Functional Genomic era. Kluwer Academic/Plenum Publishing, New York.
32. Lane, D., R. Rothenbuehler, A. M. Merrillat, and C. Aiken. 1987. Analysis of the F plasmid centromere. Mol. Gen. Genet. 207:406–412.
33. Larina, V. S. (1976). Lysogenic clones of wild-type plague bacteria and characteristics of the phages produced by them. Zh. Mikrobiol. Epidemiol. Immunobiol. 1:119–122.
34. Lobocka, M. B., A. N. Svarchevsky, V. N. Rybchin, and M. B. Yarmolinsky. 1996. Characterization of the primary immunity region of the *Escherichia coli* linear plasmid prophage N15. J. Bacteriol. 178:2902–2910.
35. Matthews, R. E. F. 1981. The classification and nomenclature of viruses. Intervirology 16:53–60.
36. Nicolle, P., H. H. Mollaret, Y. Hamon, and J. F. Vieu. 1967. Etude lysogénique, bacteriogénique, et lysotypique de l espéce *Yersinia enterocolitica*. Ann. Inst. Pasteur (Paris), 112:86–92.
37. Nicolle, P. 1973. *Y. enterocolitica*, pp. 377–387. *In* H. Rishe (ed) Lysotypie und andere spezielle, epidemiologische Laboratoriumsmethoden. Gustav Fischer , Jena.
38. Nicolle, P., H. H. Mollaret, and J. Brault. 1976. New results on the phage typing of *Yersinia enterocolitica*, concerning more of 4000 strains of various origins. Rev. Epidemiol. Sante Publ. 24:479–496.
39. Nilehn, B., and C. Ericson. 1969. Studies on *Yersinia enterocolitica* bacteriophages liberated from chloroform treated cultures. Acta Pathol. Microbiol. Scand. 75:177–187.
40. Novosel'tsev, N. N., I. V. Ryzhko, and V. K. Kidreev. 1977. Origin of plague temperate phages of serotype 2. Zh. Mikrobiol. Epidemiol. Immunobiol. 10:111–115.
41. Pajunen, M., S. Kiljunen, and M. Skurnik. 2000. Bacteriophage ϕYeO3-12 specific for *Yersinia enterocolitica* serotype O:3 is related to coliphages T3 and T7. J. Bacteriol. 182:5114–5120.

42. Pajunen, M. I., S. Kiljunen, E. L. Söderholm, and M. Skurnik. 2001. Complete genomic sequence of the lytic bacteriophage ϕYeO3–12 of *Yersinia enterocolitica* serotype O:3. J. Bacteriol. 183:1928–1937.
43. Pajunen, M. I., M. R. Elizondo, M. Skurnik, J. Kieleczawa, and I. J. Molineux. 2002. Complete nucleotide sequence and likely recombinatorial origin of bacteriophage T3. J. Mol. Biol. 319:1115–1132.
44. Pansegrau, W., and E. Lanka. 1992. A common sequence motif among prokaryotic DNA primases. Nucleic. Acids Res. 20:4931.
45. Popp, A., S. Hertwig, R. Lurz, and B. Appel. 2000. Comparative study of temperate bacteriophages isolated from *Yersinia*. Syst. Appl. Microbiol. 23:469–478.
46. Ravin, N., and D. Lane. 1999. Partition of the linear plasmid N15: interaction of N15 partition functions with the *sop* locus of the F plasmid. J. Bacteriol. 181:6898–6906.
47. Ravin, N. V., A. N. Svarchevsky, and G. Dehò. 1999. The anti–immunity system of phage-plasmid N15: identification of the antirepressor gene and its control by a small processed RNA. Mol. Microbiol. 34:980–994.
48. Ravin, N. V., T. S. Strakhova, and V. V. Kuprianov. 2001. The protelomerase of the phage-plasmid N15 is responsible for its maintenance in linear form. J. Mol. Biol. 312:899–906.
49. Ravin, N. V., V. V. Kuprianov, E. B. Gilcrease, and S. R. Casjens. 2003. Bidirectional replication from an internal *ori* site of the linear N15 plasmid prophage. Nucleic Acids Res. 31:6552–6560.
50. Ravin, V., N. Ravin, S. Casjens, M. E. Ford, G. F. Hatfull, and R. W. Hendrix. 2000. Genomic sequence and analysis of the atypical temperate bacteriophage N15. J. Mol. Biol. 299:53–73.
51. Recktenwald, J., and H. Schmidt. 2002. The nucleotide sequence of Shiga toxin (Stx) 2e-encoding phage ϕP27 is not related to other Stx phage genomes, but the modular genetic structure is conserved. Infect. Immun. 70:1896–1908.
52. Rybchin, V. N., and A. N. Svarchevsky. 1999. The plasmid prophage N15: a linear DNA with covalently closed ends. Mol. Microbiol. 33:895–903.
53. Sharp, R., I. S. Jansons, E. Gertman, and A. M. Kropinski. 1996. Genetic and sequence analysis of the *cos* region of the temperate *Pseudomonas aeruginosa* bacteriophage, D3. Gene 177:47–53.
54. Skurnik, M. 1999. Molecular genetics of *Yersinia* lipopolysaccharide, pp. 23–51. *In* J. Goldberg (ed.) Genetics of Bacterial Polysaccharides. CRC Press, Boca Raton, Fla.
55. Snellings, N. J., M. Popek, and L. E. Lindler. 2001. Complete DNA sequence of *Yersinia enterocolitica* serotyp O:8 low-calcium-response plasmid reveals a new virulence plasmid-associated replicon. Infect. Immun. 69:4627–4638.
56. Spoerel, N., P. Herrlich, and T. A. Bickle. 1979. A novel bacteriophage defence mechanism the anti-restriction protein. Nature 278:30–34.
57. Steven, A. C., B. L. Trus, J. V. Maizel, M. Unser, D. A. D. Parry, J. S. Wall, J. F. Hainfeld, and F. W. Studier. 1988. Molecular substructure of a viral receptor-recognition protein. The gp17 tail-fiber of bacteriophage T7. J. Mol. Biol. 200:351–365.
58. Tsubokura, M., K. Otsuki, and Y. Kawaoka. 1982. Lysogenicity and phage typing of *Yersinia enterocolitica* isolated in Japan. Jpn. J. Vet. Sci. 44:433–438.
59. Wolf-Watz, H., D. A. Portnoy, I. Bölin, and S. Falkow. 1985. Transfer of the virulence plasmid of *Yersinia pestis* to *Yersinia pseudotuberculosis*. Infect. Immun. 48:241–243.
60. Woods, D. E., J. A. Jeddeloh, D. L. Fritz, and D. DeShazer. 2002. *Burkholderia thailandensis* E125 harbors a temperate bacteriophage specific for *Burkholderia mallei*. J. Bacteriol. 184:4003–4017.
61. Ziegelin, G., and E. Lanka. 1995. Bacteriophage P4 DNA replication. FEMS Microbiol. Rev. 17:99–107.

35

Temperate Bacteriophages of *Bacillus subtilis*

PAMELA S. FINK
STANLEY A. ZAHLER

Bacillus subtilis strains, like many other bacteria, contain temperate bacteriophage genomes within their chromosomes. Some of these bacteriophages form viable infective particles upon induction, for example ϕ105 and SP β. Other bacteriophages, such as PBSX, appear to be remnants of intact viruses that no longer have a complete genome but package fragments of the bacterial chromosome during phage maturation.

B. subtilis temperate bacteriophages come in a variety of shapes and sizes, which will be detailed in the next section. Despite physical differences, these phages have one thing in common: only linear, double-stranded DNA, be it bacteriophage or bacterial chromosome in origin, is packaged in the phage particle. There have been no reports of single-stranded DNA phages in *Bacillus* species, and only one report of an RNA phage, AP50, which is specific for *Bacillus anthracis* (89).

A chapter on *B. subtilis* temperate bacteriophages appeared in the first edition of *The Bacteriophages* (151). Since that time, a wealth of information has been reported about these phages. New isolates have been identified, phage-specific gene products have been characterized, and the DNA sequences of the entire SPβ and PBSX genomes have been determined.

Physical Characteristics of Bacteriophages

B. subtilis temperate phages are organized into five groups based on phage immunity, serology, host range, virion and genome size, DNA homology, and restriction fragment maps (21, 143, 151, 152). Table 35-1 indicates key features that define the groups and lists well-characterized members of each group. *Bacillus* phages in groups I, II, and III, which have long noncontractile tails, belong to the viral family *Siphoviridae*. Group V phages, with contractile tails, belong to the family *Myoviridae* (88). Group V phages are distinct from other phage types in that they do not package their own genomes during lytic growth, but incorporate small fragments of chromosomal DNA into their heads (96). See chapter 2 for a general discussion of phage classification based on virion morphology.

Phage Serology, Immunity, and Host Range

Antiserum prepared against one member of a group neutralizes other phages of the same group, but not phages in different groups. For example, antiserum raised against phage ϕ105 (group I) does not react with ϕ3T (group III) but does neutralize ρ6, another group I phage. Within the group III phages, it appears that there may be more than one serological group. Antiserum raised against SPβ neutralizes SPβ itself as well as phages ϕ3T, ρ11, IG1, IG3, and IG4 but not SPR or H2 (36, 139). Anti-SPR antiserum neutralizes only SPR (92), and anti-H2 antiserum inactivates only H2 (138).

Similarly, when a host cell is lysogenized by a phage, it is immune to infection by other phages in the same group but may be infected by phages of another group. An exception to this general rule is seen with *B. subtilis* strain 168, which carries group III SPβ as an endogenous prophage. Group III phages ϕ3T and ρ11 form plaques on this strain, but phages Z, SPR, IG1, IG3, IG4, and H2 do not (21, 36). There must thus be two distinct immunity types within the group III phages. Based on host range and DNA homology studies, Weiner and Zahler (139) proposed that group III phages be subdivided into three subgroups, with SPβ, ϕ3T, ρ11, IG1, IG3, and Z in subgroup 1, SPR in subgroup 2, and H2 in subgroup 3.

B. subtilis strain 168 serves as a host for phages from each of the five groups. This strain normally carries SPβ and the defective phage PBSX as resident prophages (21, 121, 137). *B. subtilis* strain R is lysogenic for SPR, which was initially misidentified as SPβ (130). Only phage SP16 forms plaques on *B. subtilis* strain W23. In addition, this phage forms plaques on *Bacillus amyloliquefaciens* strain H, *Bacillus pumilis*, *Bacillus licheniformis*, and *Bacillus globigii*,

Table 35-1 Physical Characteristics of *B. subtilis* Bacteriophages

Group	Phage members[a]	Head size (nm)	Tail size[b] (nm)	DNA length (kb)	DNA %G+C (mol%)	References
I	**ϕ105**, ρ6, ρ10, ρ14	50–52	10 × 220 (N)	38–40	43.5	7, 13
II	**SPO2**	50	10 × 180 (N)	40	43	8, 13, 14, 105
III	**SPβ**, ϕ3T, Z, ρ11, SPR, IG1, IG3, IG4, H2	72 × 82	12 × 358 (N)	110–134	31	22, 28, 35, 36, 54, 85, 132, 137, 153, 155
IV	**SP16**	61 × 61	12 × 192	60	37.8	84, 97
V	**PBSX**, PBSW, PBSY, PBSZ	45 × 45	20 × 200 (C)	30, 13 kb bacterial DNA packaged	43[c]	2, 121, 144

[a] Typical phage for each group is indicated by bold-faced type.
[b] N, noncontractile tail; C, contractile.
[c] Mol%G+C of packaged *B. subtilis* chromosomal DNA (69).

which are close relatives of *B. subtilis* (21, 84). Phage H2 also lysogenizes *Bacillus amyloliquefaciens* strain H and *Bacillus pumilis* (138).

B. subtilis strain W23 carries defective phage PBSZ, which resembles PBSX (137). After induction and lytic growth, these defective bacteriophages bind to other nonlysogenic *Bacillus* strains and kill them. The term "phibacin" (phage-like bacteriocin) has been coined to describe the killing effect of phages such as PBSX (82).

Particle and Genome Sizes

Group I and II phages have icosahedral heads about 50 nm in diameter, whereas group III and IV phages have somewhat larger heads (see table 35.1). All these phages, with the possible exception of SP16, have long, thin, flexible noncontractile tails. Figure 35.1 is an electron micrograph of the group III phage SPβ that demonstrates the typical appearance of *B. subtilis* temperate phages. At the end of the tail there is a six-lobed foot structure that probably functions in attaching the phage particle to its receptor in the bacterial cell wall.

Additional *Bacillus* temperate phages have recently been reported by Ackermann et al. (1). One of these, phage species SN45, has an elongated head (80 × 40 nm) that encompasses its 39 kb genome. At the end of its 287 nm long tail there is a baseplate resembling that of ϕ105, leading to speculation that the two phages might be related.

The linear, double-stranded DNA genomes of the phages in groups I and II are about 40 kb long, while the SP16 genome is larger, about 60 kb. Phages in group III have genomes greater than 110 kb in length, about 2–2.5 times longer than that of bacteriophage λ. Group III phage genomes contain genes essential for phage propagation, prophage maintenance, and several additional genes with diverse functions (see below).

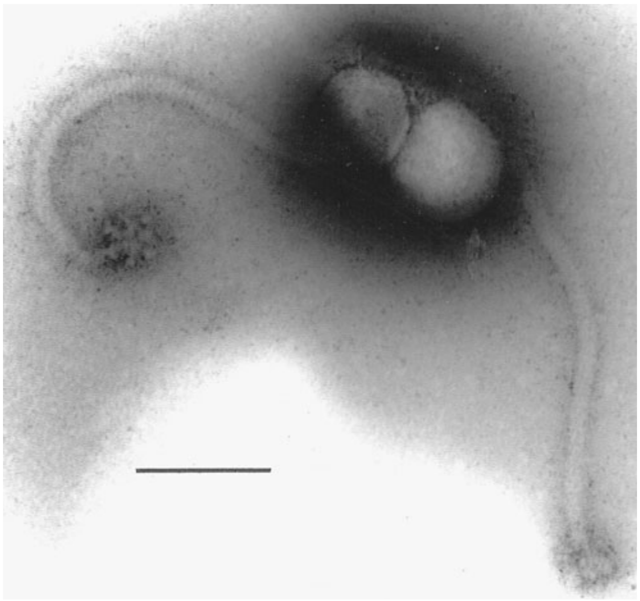


Figure 35-1 Electron micrograph of a negatively-stained preparation of SPβ particles. The long, flexible, non-contractile tail ends in a complex structure that probably is involved in phage attachment to the bacterial cell. Bar represents 100 nm.

The DNA sequence of the entire SPβ genome has been determined. It consists of 134,416 bp (74). The genome contains 187 putative open reading frames (ORFs), many of which are less than 100 amino acids long. A recently completed peptide profile of the *B. subtilis* genome revealed that small peptides, that is peptides encoded by fewer than 85 codons, are commonly organized into clusters as part of a prophage genome (156). The function of these small peptides is for the most part unknown; however, *sspC*, encoding a 72-amino acid small, acid-soluble spore protein (16) is one of the ORFs in the SPβ genome.

The ORFs are divided into clusters based on function. Cluster I [nucleotide (nt) 41–21224] comprises ORFs adjacent to the left prophage attachment site, which may be involved in phage integration and bacteriocin expression. Cluster II (nt 21280–64109) contains genes possibly associated with late phage function, such as cell lysis and perhaps phage structural proteins. Cluster III (nt 65062–134361) contains genes that act early in the phage life cycle, for example the putative repressor gene, *yonR*.

The SPβ sequencing project revealed that the %G + C content of the phage was on average 34.6 mol%, which correlates well with the previously reported value (137). Because of the variance in %G + C of phage DNA versus that of chromosomal DNA, which is on average 43.5 mol% (69), there is a notable difference in codon usage by the phage. Phage genes seem to use TTG, GTG, CTG and even ATT as start codons, as well as the common ATG. Moszer et al. (87) propose that the presence of codon bias is an indicator of systematic lateral gene transfer in *B. subtilis* that might result from prophage integration.

For the group V phages, determination of genome size is slightly more difficult because these phages package chromosomal DNA fragments, not their own genomes. Although PBSX packages 13 kb of chromosomal DNA (2), Wood et al. (144) determined that the PBSX genome itself is about 30 kb long by sequencing chromosomal DNA encompassing the PBSX prophage. The genome comprises at least a repressor gene (*xre*), phage head and tail proteins, and cell lysis functions (see below).

A phage resembling PBSX, PBND8, has been isolated from the related bacterium, *Bacillus natto* (131). The overall phage structure is similar but PBND8 has a smaller head containing a correspondingly shorter length of chromosomal DNA, only about 8 kb. It appears that the group V phages package specific lengths of chromosomal DNA based on the size of the phage head.

The head sizes of phages in groups I–IV accommodate a phage genome length of DNA. Occasionally, ϕ105 packages about 1 kb more than its genome length of 39.2 kb (32) without adverse effects on the phage head. Sometimes phages carry DNA molecules that have lost between 2 and 14 kb from their genomes. Viable deletion mutants have been identified for phages ϕ105 (43), SPO2 (50), ρ11 (63), ρ14 (68), and SPβ (39, 41, 118). These deletion mutants define specific regions not essential for phage propagation or lysogeny, providing useful information for the construction of phage cloning vectors (see below).

The ends of the linear phage DNA molecules examined thus far are of two varieties. SP16 contains terminally redundant, circularly permuted double-stranded ends (97). This suggests that the phage packages its DNA using a headful mechanism like that of *Salmonella* phage P22 (113) (reviewed in chapter 29). Other phages, including ϕ105 (111), and SP02 (13), have constant cohesive ends resembling those of phage λ (see chapter 27) except that the *Bacillus* phage DNAs have 3′ single-stranded extensions rather than 5′. This type of specific single-stranded extension is probably generated, as it is in λ, by a precise endonucleolytic cleavage that occurs during phage packaging (11). The ends of the ϕ105 genome have the sequences 5′-GCGCTCC-3′ and 3′-CGCGAGG-5′ (29). Regardless of the end structure of the phage genome, these ends must join to convert the linear molecule into a circular one prior to prophage integration.

Establishing and Maintaining Lysogeny

Adsorption to Cells

To initiate infection, *B. subtilis* temperate phages first bind to the bacterial cell. Proteins at the end of the phage tail adsorb to specific receptors on the surface of the host cell. Phage-resistant *B. subtilis* mutants have been isolated; most of the defects map to the *gtaB–tagB* region of the chromosome, which contains genes involved in teichoic acid synthesis. One such mutant, *pha-3*, eliminates adsorption by all group III phages except SPR (33). *gtaB* encodes UDP-glucose pyrophosphorylase, which functions to glycosylate cell wall teichoic acids (136). Different mutations in *gtaB* alter or eliminate bacteriophage ϕ3T adsorption (116). Another locus, *gneA*, located between *sacA* and *purA*, encodes UDP-*N*-acetylglucosamine 4-epimerase (34). Mutations in this gene lead to cells with galactosamine-deficient teichoic acid, and these cells are also resistant to ϕ3T adsorption.

The defective *B. subtilis* phages also bind to teichoic acid in the cell walls, but with different specificities. PBSX does not bind to host *B. subtilis* strain 168 cell walls, but will attach to non-host strain W23 cell walls to initiate killing of the non-host strain. Conversely, phage PBSZ will only adsorb to non-host strain 168 cells and not to host strain W23 cells. Karamata and coworkers (61, 149) found that the *tag*1 gene in strain 168, mediating glycerol teichoic acid synthesis, can be replaced by the *tar* gene from strain W23, which specifies ribitol teichoic acid synthesis. Such interstrain hybrids show the phage sensitivity pattern of strain W23, indicating that the specificity of phage binding depends on whether the cells have ribitol or glycerol teichoic acid in their walls.

Attachment Sites

Bacillus phage genomes insert into the bacterial chromosome by a Campbell-type, single crossover event (10) (see also chapter 7). The linear phage DNA circularizes, then integrates at its specific attachment site, *attB* (140). The known chromosomal locations of phage integration sites are diagrammed in figure 35.2. The ϕ105, SP02, PBSX, and H2 attachment sites occur at different locations in the chromosome (20, 47, 58, 108, 115, 144, 153). In contrast, the attachment sites for phages ϕ3T, SPβ, IG1, IG3, and IG4 are clustered between *ilvA* and *gltA*, near the terminus of DNA replication (37, 56, 142, 154). These different attachment sites may represent remnants of ancestral prophage DNA (103).

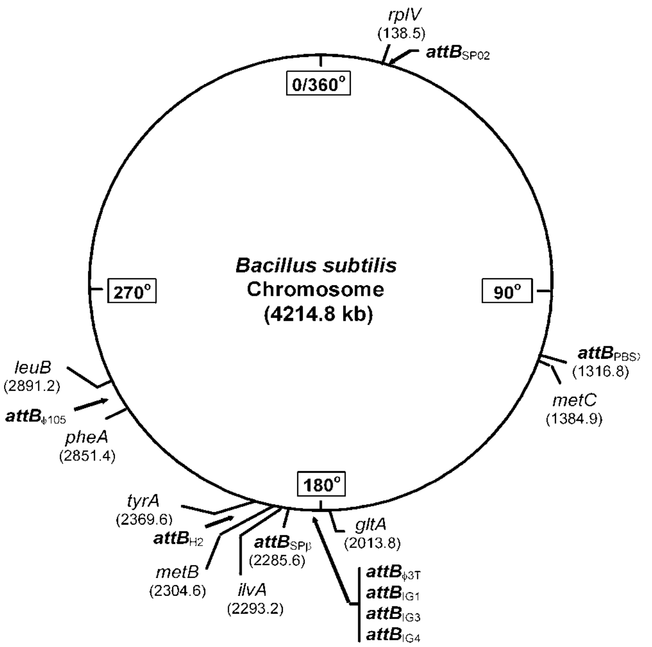

rplV
(138.5)
attB SP02
0/360°
Bacillus subtilis
Chromosome
(4214.8 kb)
270°
90°
attB PBSX
(1316.8)
metC
(1384.9)
leuB
(2891.2)
attBϕ105
pheA
(2851.4)
180°
tyrA
(2369.6)
attB H2
attB SPβ
(2285.6)
gltA
(2013.8)
attBϕ3T
attB IG1
attB IG3
attB IG4
metB
(2304.6)
ilvA
(2293.2)

Figure 35-2 Location of bacteriophage attachment sites in the *B. subtilis* chromosome. Genetic landmarks and approximate location of prophage attachment sites are indicated. Numbers refer to kilobase position in *B. subtilis* genome (69).

B. subtilis strain 168 contains SPβ and PBSX as endogenous prophages, but strains have been constructed that lack both prophages. These cured strains recombine, repair damaged DNA and sporulate as well as the lysogenized strain (148). In addition, despite the fact that SPβ lies near the terminus of chromosomal DNA replication, elimination of the SPβ prophage does not alter where chromosomal DNA replication ends (59).

For phage SPβ, the bacterial (*attB*), phage (*attP*), and prophage (*attL* and *attR*) attachment-site sequences have been identified (12, 74, 146); they are diagrammed in figure 35-3. The SPβ attachment sites resemble those identified for coliphage λ. The λ attachment sites contains an AT-rich core of identical nucleotides, and the λ *attP* core is flanked by regions of high similarity (140). The SPβ core region is also AT-rich (30%G + C), having the nucleotide sequence 5′-ATACAGCTTTATCTGT-3′. Segments flanking the SPβ *attP* core contain inverted repeats (figure 35-3, arrows) that resemble integrase binding sites in phage λ DNA (107). There are also AT-rich regions in *attP*, which might correspond to integration host factor protein (IHF) binding sites (104). IHF binding at phage λ *attP* results in DNA bending, facilitating site-specific recombination mediated by λ integrase (86).

An integration-deficient variant of phage SPβ, the *int-5* mutant, has been isolated (150). This phage rarely inserts at the normal SPβ *attB* adjacent to *ilvA*, but inserts at other sites in the *B. subtilis* genome. These alternative insertions may result from recombination between SPβ DNA and regions of chromosomal DNA that have strong homology to the phage genome (74, 103). Apparently the lack of integrase function also affects excision of the SPβ *int-5* mutants. These mutants are not efficiently induced, and when they do excise from the chromosome, they delete bacterial genes adjacent to their insertion sites. For example, excision of the SPβ *int-5* phage integrated at its normal attachment site can result in the loss of DNA from the prophage to beyond *terC*, a deletion of about 230 kb (57).

Phage Repressor and Genetic Controls

To maintain lysogeny, prophage gene expression must be controlled by a repressor protein. ϕ105, SPβ, ρ6, and

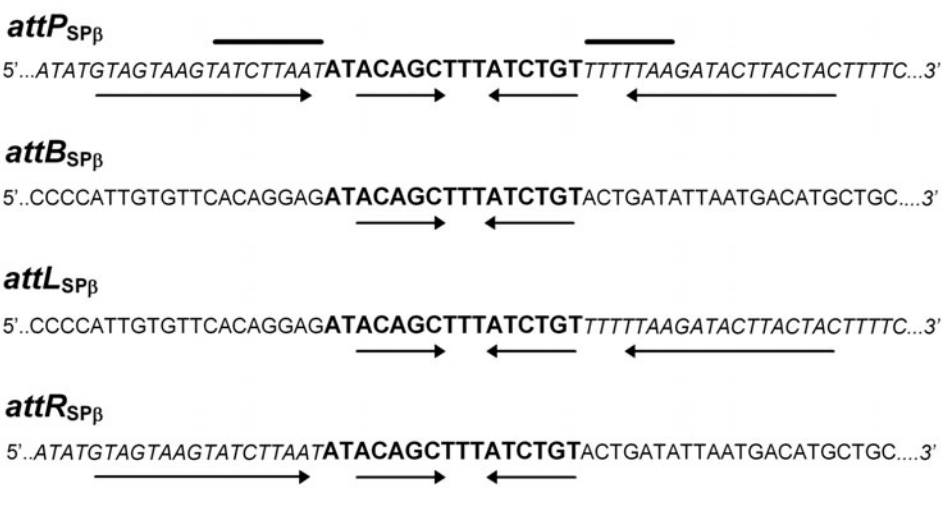

attP SPβ
5′...ATATGTAGTAAGTATCTTAATATACAGCTTTATCTGTTTTTTAAGATACTTACTACTTTTC...3′
attB SPβ
5′..CCCCATTGTGTTCACAGGAGATACAGCTTTATCTGTACTGATATTAATGACATGCTGC....3′
attL SPβ
5′..CCCCATTGTGTTCACAGGAGATACAGCTTTATCTGTTTTTTAAGATACTTACTACTTTTC...3′
attR SPβ
5′..ATATGTAGTAAGTATCTTAATATACAGCTTTATCTGTACTGATATTAATGACATGCTGC....3′

Figure 35-3 DNA sequence of SPβ attachment sites. Attachment sites are those associated with the bacteriophage (*attP*$_{\mathrm{SP}\beta}$) the bacterial chromosome (*attB*$_{\mathrm{sP}\beta}$) and the prophage (*attL*$_{\mathrm{sP}\beta}$ and *attR*$_{\mathrm{sP}\beta}$). Phage DNA sequences are italicized. Core region is in bold-face type, arrows denote inverted repeats, AT-rich regions overlined.

φ3T phages carrying defective repressor (*c*) genes have been isolated (18, 21, 137). With a nonfunctional repressor, these phage cannot lysogenize their hosts, yielding a clear-plaque morphology.

The gene encoding the 144 amino acid φ105 repressor protein, $c_{\phi105}$, has been cloned and sequenced (19, 25). The N-terminal end of the repressor contains a helix–turn–helix (HTH) motif similar to the λ cro repressor (133). φ105 mutants with alterations occurring within the first 43 residues of the protein cannot bind to phage promoter regions. The repressor functions as a tetramer, binding to operator sites in the φ105 immunity region (*immF*).

The organization of the φ105 genetic control region is diagrammed in figure 35-4A. As with phage λ repressor, the φ105 repressor interacts with two divergent promoters in the immunity region. Transcription from P_R leads to synthesis of late phage genes and lytic growth, while transcription from P_M leads to $c_{\phi105}$ synthesis and lysogeny. By constructing fusions between these promoters and the *cat*-86 gene, Van Kaer and coworkers (134) demonstrated that the φ105 repressor stimulates expression from P_M and represses expression from P_R. The repressor protein binds to six operator (O_R) sites, three of which (O_R1, O_R2, and O_R3) have a common sequence, 5′-GACGGAAATACAAG-3′, which is unusual in that it does not show 2-fold rotational symmetry (134, 135). O_R4 and O_R5 differ from this consensus sequence in two residues and O_R6 differs at five sites. All the operator sequences lie within the approximately 200 bp region between the two promoters with the exception of O_R3, which occurs about 250 bp downstream of P_R. The φ105 repressor binds tightly to O_R1, O_R2, O_R3, and O_R6; binding of the repressor at these sites precludes transcription from P_R. Weaker binding occurs at O_R4 and O_R5; repressor binding at these sites may autoregulate repressor expression.

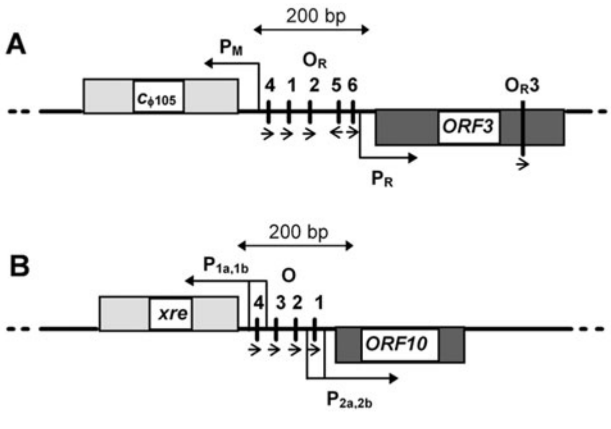


Figure 35-4 Organization of genetic control regions. A: Phage φ105. B: Phage PBSX. Small arrows indicate orientation of operator sequences. For further details, see (135) and (81) for φ105 and PBSX, respectively.

The PBSX repressor gene, *xre*, has also been cloned and sequenced (145). The repressor (Xre) is 113 amino acids long and contains an HTH motif at its N-terminal end. A temperature-sensitive repressor mutant, *xhi*1479 (9), differs from the wild-type protein at three residues, one of which is in the HTH region. The promoter region of the *xre* gene is diagrammed in figure 35-4B. Another gene, encoding ORF10, is adjacent to *xre* but divergently transcribed, which resembles the situation in the φ105 *immF*. Between *xre* and ORF10 there are four 15 bp palindromic sequences, 5′-GATACATTTTGTATC-3′, which bind to purified Xre protein (81, 82). These sites have been called O1, O2, O3, and O4. At low concentrations, Xre binds first to O1 and O2 to prevent transcription rightward through ORF10 and the late genes, but at higher concentrations it can prevent its own synthesis by binding at O3 and O4 (81).

In the SPβ genome, the *yonR* ORF has been assigned as the repressor gene by virtue of its homology with the PBSX *xre* gene (74). Another gene involved in SPβ immunity, *d* or *yomJ*, was previously cloned and shown to confer SPβ immunity to the host cell (80). However, the *d* gene does not complement SPβ *c* mutants; thus it is not the main phage repressor gene. A similar situation occurs in φ105, where genes on two different cloned DNA fragments can confer phage immunity to the host, but only one of the genes actually produces a protein that binds to φ105 operator sequences (24).

Prophage Induction and Specialized Transduction

Prophage Induction

All *B. subtilis* temperate phages are induced to enter lytic growth when the host cell DNA has been damaged. Physical and chemical treatments, including mitomycin C, UV light and *N*-methyl-*N*-nitrosoguanidine (NTG), stimulate the cellular SOS-like response, which has been called the SOB regulon in *B. subtilis* (147). Expression of the *B. subtilis* RecA protein is stimulated following DNA damage. In the presence of RecA, prophage repressor proteins are inactivated, resulting in phage induction. The phage genome excises from the chromosome by a reversal of the site-specific recombination event that led to phage integration (for review of related regulatory circuitry as displayed by phage λ, see chapter 8).

Some phages are induced by other cell-damaging treatments, such as exposure to hydrogen peroxide, stimulates growth of PBSX and possibly SPβ (53, 122). In addition, IG1 is induced by the DNA polymerase III inhibitor, 6-(*p*-hydroxyphenylazo)-uracil, and can multiply effectively despite a drastic decline in host cell viability (38). When *B. subtilis* cells develop competence, some of the SOB-related genes, such as *recA*, are activated. Prophages of φ105 and

SP02 are induced during competence, but ϕ3T and SPβ prophages have developed a mechanism to prevent induction during competence development (83).

Temperature-sensitive repressor mutations have also been isolated for ϕ105 [*cts*23, (3)], SPβ, [*c*2, (106)], and PBSX [*xhi*1479 (9, 145)]. Following a brief heat shock (48–50°C), the repressor function is destroyed and the prophage enters lytic growth.

During sporulation, the *skin* (sigK intervening) element, a 42 kb long segment of the chromosome at position 230°, is released by a site-specific recombination event resembling prophage excision (124). DNA sequence analysis of the *skin* element reveals strong homology between it and PBSX, including genes for a potential repressor and an autolysin (67). Despite the high degree of similarity between the two DNAs, the *skin* late gene operon is not expressed and no phage particles are produced. The *skin* element may represent a remnant of some ancestral prophage.

Specialized Transducing Phages

Occasionally prophage excision is not precise and recombination takes place between phage DNA and adjacent chromosomal DNA, which results in the formation of specialized transducing particles (150). ϕ105 mediates specialized transduction of genes flanking its attachment site, that is *ilvBC-leu* and *pheA*, but at low frequency and only by replacement of defective alleles in recipient cells (114).

In contrast, SPβ mediates specialized transduction of genes flanking its normal attachment site, that is *ilvD-thyB-ilvA* and *kauA-odh (citK)*, but it generates high frequency of transduction (HFT) lysates (40, 106, 154). These HFT lysates have transducing particle-to-phage particle ratios between $1{:}10^2$ and $1{:}10^4$. When SPβ integrates at secondary sites in the chromosome, it also mediates specialized transduction of genes flanking the prophage, for example *ilvBC-leu*, *dad-ddl*, *glnA*, and *degU* (46, 77, 79, 100). Phages IG1, IG3, IG4, H2, and ϕ3T also mediate specialized transduction of genes adjacent to their chromosomal attachment sites (37, 95, 153).

Only one nondefective specialized transducing particle has been reported SPβ:SPβ *c*2 p*ilvA* (42). Figure 35-5 depicts how this specialized transducing phage may have been generated. DNA sequence data obtained from the SPβ *c*2 p*ilvA* phage (4) have been compared with the DNA upstream of *ilvA* and to SPβ sequences (69, 74). Apparently there is a region of approximately 27 bp in the bacterial chromosome upstream of *ilvA* [nt 2293436–2293462 (69)] that is very homologous to a part of SPβ near the right end of the prophage [nt 126571–126588 (74)], with 22 of 27 bases identical. Possibly this short region of homology was enough to serve as a recombination site during the genesis of the SPβ *c*2 p*ilvA* phage. This finding lends support to the notion that ancestral phage DNA exists in the chromosome (103). In the SPβ *c*2 p*ilvA* particle, approximately 8 kb of the phage genome are replaced by DNA from the bacterial chromosome; prophage genes to the right of *yosR* do not appear to encode essential phage proteins.

Phage-Specific Genes

DNA Methyltransferases

The genomes of group III phages ϕ3T, SPβ, ρ11, SPR, and H2 contain methyltranferase encoding genes, which are expressed during vegetative growth of the phages. Phage-derived methyltransferase (MTase) activity was first observed in phages ϕ3T and SPR, when it was discovered

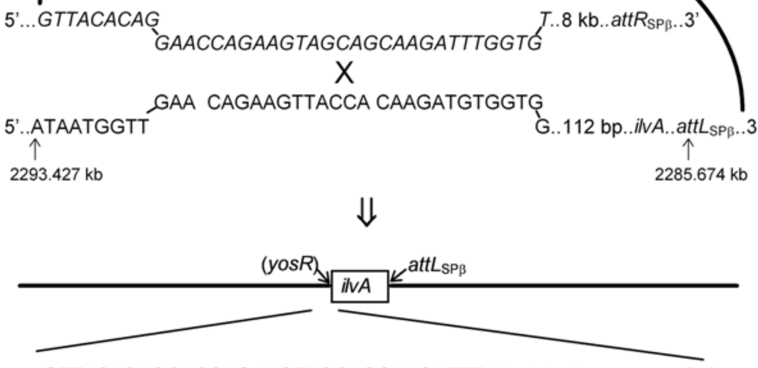

5'...GTTACACAG
GAACCAGAAGTAGCAGCAAGATTTGGTG
T..8 kb..attR SPβ..3'
X
GAA CAGAAGTTACCA CAAGATGTGGTG
5'..ATAATGGTT
G..112 bp..ilvA..attL SPβ..3'
2293.427 kb
2285.674 kb
⇓
(yosR)
ilvA
attL SPβ
5'...GTTACACAGGACCAGAAGTAGCAGCAAGATTTGGTGGACA...109 bp..ilvA...3'

Figure 35-5 DNA content of specialized transducing bacteriophage SPβ *c*2 p*ilvA*. Phage DNA sequences are italicized. Indicated kilobases denote locations in the *B. subtilis* chromosome (69).

that the phage DNAs contained *Hae*III (*Bsu*RI) restriction sites that were methylated during phage growth (17, 130). MTase modification may serve to protect the phage genomes from DNA restriction as they pass into new hosts.

The phage-encoded MTases are all type II, requiring *S*-adenosylmethionine as the methyl group donor (see 93 for a review), but some of these phage MTases are unique in that they modify multiple recognition sites. The phage MTases mediate 5-C methylation of a C residue in their respective target sequences. Most of the phage MTases modify the *Hae*III (*Bsu*RI) site, GGCC, methylating the internal C residue (92). In addition, the SPβ, ϕ3T, and $\rho 11_B$ MTases modify the *Fnu*4HI site, GCNGC, again at the internal C residue (70, 128). The three genes encoding these MTases have been isolated and expressed in *Escherichia coli*, and the proteins are all very similar in size (~47 kDa) and structural relatedness (91). The MTase from a ρ11 variant, $\rho 11_s$ modifies the *Hae*III sequence, but also acts on the *Bsp*1286 site, G (A/G/T)GC(T/C/A)C; modification of this sequence has not been completely characterized (6).

The SPR MTase has a more complex nature because it modifies three sites: *Hae*III, *Hpa*II (CCGG), and *Eco*RII (CCA/TGG). It may modify both cytosine residues, or only the internal cytosine of the *Hae*III sequence (52, 129). The H2 MTase also has triple sequence specificity, methylating the *Hae*III, *Fnu*4HI, and *Bsp*1286 sites (71). In addition, H2 carries a separate, second MTase gene that protects *Bam*HI sites (GGATCC) in the phage, modifying the internal C residue (15, 155). Similarly, Trautner and coworkers have found that ϕ3T and $\rho 11_S$ also carry a second MTase; this one recognizes the sequence TCGA, and methylates the C residue (94). Although phage Z does not inherently contain a MTase gene, it does have homology with DNA that flanks MTases in related group III phages, and a functional MTase can be recombined into the Z genome from one of these other phages (126).

Regardless of the specificity of the DNA binding sites, there are several features common to the phage MTases. The N-terminal end of the MTases is the site of *S*-adenosylmethionine binding and possibly the catalytic site, while the C-terminal end is required for general DNA interaction. Between the conserved elements, there is a variable region responsible for binding to the specific target site in the DNA (51, 94). MTases with multiple sequence specificities contain independent target-recognizing domains in a modular organization (70, 141). Evidence for the domain structure comes from isolation of mutant proteins that have lost one or more of the methylating activities, while the remaining methylase functions are normal (52). In addition, chimeric MTases can be constructed from two different MTases by "domain swapping" the target-specifying regions of the genes (5).

Thymidylate Synthetase

Most of the group III phages also carry the *thyP* gene, which encodes thymidylate synthetase (TSase), near the center of the phage genome (110, 117). The *thyP3* of ϕ3T has been cloned, and the gene complements Thy auxotrophy in both *E. coli* and *B. subtilis* (28). When the cloned *thyP3* gene is transformed into *B. subtilis* 168, the DNA recombines at one of two chromosomal locations to yield Thy^+ prototrophs (123). One site is in the *thyA* gene, which is not surprising given the high degree of sequence similarity (97%) between the *thyP3* and *thyA* genes (64, 125). *B. subtilis*, unlike any other bacteria, has a second TSase gene, *thyB*, mapping near the SPβ prophage and encoding a thermolabile TSase (90). However, the *thyP3* and *thyB* genes are less than 30% identical in their DNA sequences (60), and the *thyP* gene does not recombine in this region.

The other site of *thyP* insertion is in the resident SPβ prophage. SPβ itself does not carry a *thyP* gene, but DNA sequences flanking the ϕ3T *thyP* are homologous to a region in the center of the SPβ genome (117). Hybrid SPβ can be constructed that carry the *thyP* gene, the so-called SPβT phages (119). It is not clear why group III phages harbor the *thyP* gene, but the significant sequence similarity between *thyP* and *thyA* indicates a possible evolutionary link between these two TSases.

Lytic Proteins

During the course of lytic growth, the host cell is lysed by phage-encoded enzymes to release the newly synthesized phage. Four such lytic protein genes have been identified in the PBSX late gene operon: *xepA*, *xhlA*, *xhlB*, and *xlyA* (66, 78). The *xepA* gene encodes an exoprotein whose precise function is not yet known. *xlyA* encodes a 32 kDa endolysin resembling the *cwlA*-encoded bacterial amidase in *B. subtilis* (44). Degradation of cell walls by the PBSX amidase produces *N*-2,3-dinitrophenyl-L-alanine, indicating that it is an *N*-acetylmuramoyl-L-alanine amidase. It is the major enzyme for lysing the cell wall following PBSX induction. *xhlA* and *xhlB* encode two polypeptides (89 and 87 amino acids, respectively) that possibly interact to form a holin, which may play a role in exporting the amidase (see chapter 10 for review of holin and amidase functions).

SPβ also produces an *N*-acetylmuramoyl-L-alanine amidase, which is 40 kDa in size and is the product of the *blyA* gene (102). *blyA* is part of an operon that also contains *bhlA* and *bhlB*, which encode holin-like polypeptides of 70 and 88 amino acids, respectively. This is the same gene organization seen in PBSX and may be an example of horizontal gene exchange that might occur during phage evolution (55).

A mutant of ϕ105, ϕ105MU331, is a prophage vector that cannot lyse host cells (76). The phage contains a *lacZ* reporter gene inserted into a putative holin gene. Without holin activity, the lytic enzyme is nonfunctional.

Phages also encode other proteins that kill sensitive bacteria, such as betacin (54). The genes encoding betacin (*bet*) and tolerance to betacin (*tol*) are both present on the SPβ genome, possibly encoded by *yolG* and *yolH*, respectively (74). Although the betacin protein has not been purified, the putative protein deduced from the *yolG* sequence is about 6 kDa, and the putative *tol* gene product resembles a protein ABC-transporter.

PBSX functions as a phage-like bacteriocin, infecting and killing phage-sensitive cells. A phage-encoded protein, perhaps a tail protein, causes disruption of the cell wall and leads to lysis (120).

Other Phage Genes

Phages carry a variety of genes involved in DNA replication or recombination. SPO2 and IG1 encode their own DNA polymerases that, unlike the *B. subtilis* DNA polymerase III, are not inhibited by 6-(*p*-hydroxyphenylazo)-uracil (38, 109). SPO2 also contains a DNA repair gene(s), allowing the phage to repair UV-induced damage even in a Uvr^- host (45).

As additional DNA sequence information becomes available, more phage-associated genes are identified. The SPβ *yoqV* gene encodes an ATP-dependent DNA ligase, which is common to other phages. However, the YoqV ligase will not complement the activity of the essential NAD-dependent ligase encoded by *yerG*, and is not essential to the cell because SPβ nonlysogens are quite viable (99). Another SPβ gene, *yopP*, is homologous to the bacterial *codV*, which encodes a recombinase that resolves chromosomal dimers at the *dif* site during cell division (112). In a Cod^- SPβ lysogen, the phage recombinase reduces the frequency of defective chromosomal partitioning events.

SPβ also encodes a ribonucleotide reductase (RR), which is responsible for the reduction of ribonucleotides to deoxyribonucleotides. The two subunits of the enzyme are produced by the phage *bnrdE* and *bnrdF* genes, which are highly homologous to their *B. subtilis* counterparts, *nrdE* and *nrdF* (75). The phage genes differ from the cellular homologs in one crucial respect: each of the phage genes contains a group 1 intron, and *bnrdE* also contains an intein (48, 75). Although examples of introns, and more recently inteins, are well known in bacteria, archaea and eukaryotes, the intron–intein coincidence has not been observed before (23). In addition, Lazarevic reports that sequences in the approximately 330 bp intergenic region between *bnrdE* and *bnrdF* are similar to a eukaryotic splicesosome and may represent the remnants of another intron (72). Figure 35-6 depicts the introns, intein, and possible splicesosome-like structures.

Within the *bnrdF* intron there is a 522 bp ORF, *yosQ*, resembling an intron homing endonuclease, which is involved in the spread of introns between genes. The *bnrdE* intron does not contain such an ORF. However, the *bnrdE* intein has all the characteristic features of a homing endonuclease (75). In other RR genes from SPβ-related phages in different *Bacillus* species, the same intron and intron–intein insertions occur, and a putative homing endonuclease also appears in the *bnrdE–bnrdF* intergenic region (73). A recent review article presents a detailed discussion of intron occurrence in bacterial and phage genes (27).

Temperate Phage as Cloning Vectors

Several cloning vectors have been derived from *B. subtilis* temperate phages, including ϕ105, SPβ, ρ11, and ρ14 (30, 31, 62, 65, 68). These vectors generate stable, single-copy insertions of the cloned genes into the host chromosome. Two different methods have been developed for using phage vectors. The first is direct transfection of protoplasts by recombinant phage DNA, which has been quite successful with ϕ105 (see figure 35-7A).

The vector shown in figure 35-7A, ϕ105J106, has unique *Bam*HI, *Xba*I, and *Sal*I restriction sites for cloning, but all essential phage genes are intact (26). Following digestion with *Sal*I, the 4-base overhang (5′-TCGA. . .) is partially filled in with bases T and C to prevent self-ligation of vector "arms." Similarly, insert DNA partially digested with *Mbo*I has its 4-base overhangs (5′-GATC. . .) partially filled in with bases G and A. When vector and insert DNA are mixed, the remaining 2-base overhangs are compatible and the recombined molecules ligate together. Usefulness of this vector for "shotgun" cloning is limited because only about 5–6 kb of DNA can be inserted into the phage.

ϕ105 vectors have been constructed that promote efficient expression of specific recombinant genes (49, 127). These vectors contain a temperature-sensitive repressor protein, *cts52*, and have foreign gene expression under the control of a phage late gene promoter. One of these vectors, ϕ105MU331, has the site of foreign gene insertion in a putative holin gene (76). Following heat induction, the heterologous gene is highly expressed from the phage promoter and because the host cells do not lyse, more of the protein can be produced.

The second cloning method used is prophage transformation, whereby ligated phage and chromosomal DNA are introduced into competent lysogenic cells. The ligated DNA recombines into the prophage to generate specialized transducing phage (see figure 35-7B). This method has been used more successfully with ρ11 and SPβ than ϕ105 vectors because the latter phage is induced during competence development (see above). The SPβ vector shown in figure 35-7B, SPβ *c*2 Δ2::Tn*917*, is heat inducible, has a 10 kb deletion (Δ2), and carries the erythromycin-resistance transposon, Tn*917* (101). A plasmid vector, pCV1, carries the ends of Tn*917* and SPβ DNA flanking those

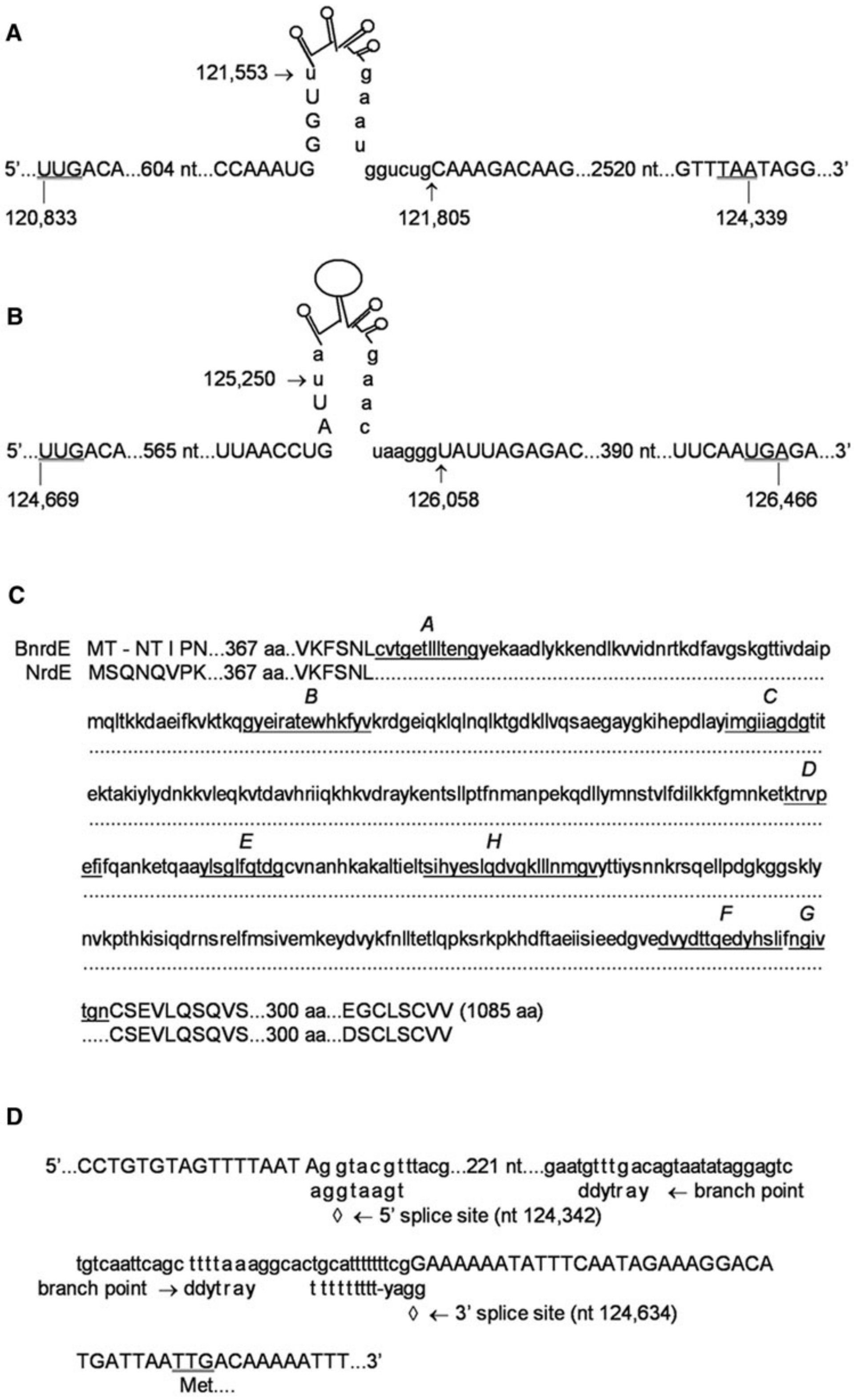

A
121,553 → u U G G
g a a u
5'...UUGACA...604 nt...CCAAAUG ggucugCAAAGACAAG...2520 nt...GTTTAATAGG...3'
120,833
121,805
124,339
B
125,250 → a u U A
g a a c
5'...UUGACA...565 nt...UUAACCUG uaagggUAUUAGAGAC...390 nt...UUCAAUGAGA...3'
124,669
126,058
126,466
C
BnrdE MT - NT I PN...367 aa..VKFSNLcvtgetllltengyekaadlykkendlkvvidnrtkdfavgskgttivdaip
NrdE MSQNQVPK...367 aa..VKFSNL..........
A
B
C
mqltkkdaeifkvktkqgyeiratewhkfyvkrdgeiqklqlnqlktgdkllvqsaegaygkihepdlayimgiiagdgtit
D
ektakiylydnkkvleqkvtdavhriiqkhkvdraykentsllptfnmanpekqdllymnstvlfdilkkfgmnketktrvp
E
H
efifqanketqaaylsglfqtdgcvnanhkakaltieltsihyeslqdvqklllnmgvyttiysnnkrsqellpdgkggskly
F
G
nvkpthkisiqdrnsrelfmsivemkeydvykfnlltetlqpksrkpkhdftaeiisieedgvedvydttqedyhslifngiv
tgnCSEVLQSQVS...300 aa...EGCLSCVV (1085 aa)
.....CSEVLQSQVS...300 aa...DSCLSCVV
D
5'...CCTGTGTAGTTTTAAT Ag gtacgtttacg...221 nt....gaatgtttgacagtaatataggagtc
aggtaagt
ddytray ← branch point
◊ ← 5' splice site (nt 124,342)
tgtcaattcagc ttttaaaggcactgcattttttcgGAAAAAATATTTCAATAGAAAGGACA
branch point → ddytray
tttttttt-yagg
◊ ← 3' splice site (nt 124,634)
TGATTAATTGACAAAAATTT...3'
Met....

Figure 35-6 Introns, intein, and possible splicesosome in the *bnrd* region of SPβ. A: *bnrdE* intron, B: *bnrdF* intron, C: BnrdE intein, and D: possible intergenic splicesomal element. Numbers indicate nucleotide position in SPβ *c2* prophage sequence (74), underlined sequences in C are conserved intein motifs (75), and double-underlined sequences are start/stop codons.

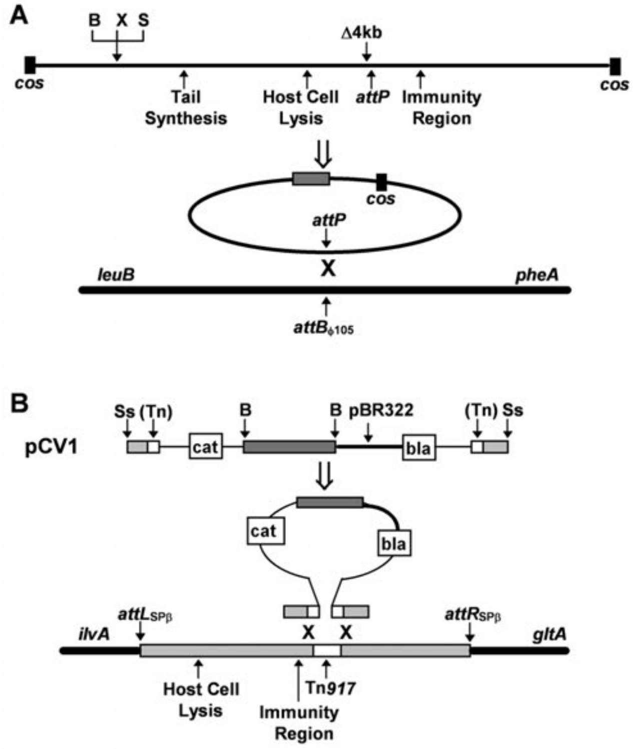


Figure 35-7 Cloning vectors constructed from phages. A: Direct transfection with phage ϕ105J106 and B: prophage transformation with phage SPβ *c*2 Δ2::Tn*917*. Dark-gray bars represent insert DNA; light-gray bars represent SPβ DNA; B is *Bam*HI; X is *Xba*I; S is *Sal*I; Ss is *Sst*I.

ends on a pBR322-derived plasmid vector that replicates only in *E. coli*. The plasmid is cleaved into two fragments by *Sst*I and *Bam*HI, heterologous DNA is inserted into the *Bam*HI site, and long concatemers of vector–insert–vector DNAs are generated during ligation. The concatemers are transformed into the SPβ *c*2 Δ2::Tn917 lysogen and, by homologous recombination, the cloned DNA is integrated into the prophage to generate a specialized transducing phage. The recombinant phages are identified by selection for chloramphenicol resistance, and these drug-resistant clones are pooled and subjected to heat induction. To isolate clones carrying specific genes, the recombinant phage lysates are used to convert an auxotrophic nonlysogen to prototrophy.

Conclusions

Despite the diverse nature and unique features of the *B. subtilis* temperate phages, they have similarities to temperate phage isolated from another bacteria, for example in their genetic control. Some of their functions resemble those of *B. subtilis* lytic phage, such as their cell-lysis mechanisms (98). It is becoming more evident that these conserved mechanisms are no accident, but have resulted from a high degree of horizontal gene transfer between the phage genomes (56).

Acknowledgements

This work is dedicated to the memory of Dr. Ruth Z. Korman, whose passion for life and science were an inspiration.

References

1. Ackermann, H.-W., R. R. Azizbekyan, H. P. Emadi Konjin, M.-M. Lecadet, L. Seldin, and M. X. Yu. 1994. New *Bacillus* bacteriophage species. Arch. Virol. 135:333–344.
2. Anderson, L. M., and K. F. Bott. 1985. DNA packaging by the *Bacillus subtilis* defective bacteriophage PBSX. J. Virol. 54:773–780.
3. Armentrout, R. W., and L. Rutberg. 1971. Heat induction of ϕ105 in *Bacillus subtilis*: replication of the bacterial and bacteriophage genomes. J. Virol 8:455–468.
4. Armpriester, J., and P. S. Fink. 1990. DNA sequence analysis of the *ilvA* gene from *Bacillus subtilis* 168. Abstracts of the 90th General Meeting of the American Society for Microbiology, p. 89.
5. Balganesh, T. S., L. Reiners, R. Lauster, M. Noyer-Weidner, K. Wilke, and T. A. Trautner. 1987. Construction and use of chimeric SPR/ϕ3T DNA methyltransferases in the definition of sequence recognizing enzyme regions. EMBO J. 6:3543–3549.
6. Behrens, B., M. Noyer-Weidner, B. Pawlek, R. Lauster, T. S. Balganesh, and T. A. Trautner. 1987. Organization of multispecific DNA methyltransferases encoded by temperate *Bacillus subtilis* phages. EMBO J. 6:1137–1142.
7. Birdsell, D. C., G. M. Hathaway, and L. Rutberg. 1969. Characterization of temperate *Bacillus* bacteriophage ϕ105. J. Virol. 4:264–270.
8. Boice, L. B. 1969. Evidence that *Bacillus subtilis* bacteriophage SP02 is temperate and heteroimmune to bacteriophage ϕ105. J. Virol. 4:47–49.
9. Buxton, R. S. 1976. Prophage mutation causing heat inducibility of defective *Bacillus subtilis* bacteriophage PBSX. J. Virol. 20:22–28.
10. Campbell, A. 1962. Episomes. Adv. Genet. 11:101–145.
11. Catalano, C. E., D. Cue, and M. Feiss. 1995. Virus DNA packaging: the strategy used by phage lambda. Mol. Microbiol. 16:1075–1086.
12. Chang, C. 1991. Characterization of the bacterial attachment site for *Bacillus subtilis* bacteriophage SPβ, MS thesis, Wright State University, Dayton, Ohio.
13. Chow, L. T., L. Boice, and N. Davidson. 1972. Map of the partial sequence homology between DNA molecules of *Bacillus subtilis* prophages SPO2 and ϕ105. J. Mol. Biol. 68:391–400.

14. Chow, L. T., and N. Davidson. 1973. Electron microscope study of the structures of the *Bacillus subtilis* prophages, SPO2 and φ105. J. Mol. Biol. 75:257–264.
15. Connaughton, J. F., W. D. Kaloss, P. G. Vanek, G. A. Nardone, and J. G. Chirikjian. 1990. The complete sequence of the *Bacillus amyloliquefaciens* proviral H2, *Bam*HI methylase gene. Nucleic Acids Res. 18:4002.
16. Conners, M. J., and P. Setlow. 1985. Cloning of a small, acid-soluble spore protein gene from *Bacillus subtilis* and determination of its complete nucleotide sequence. J. Bacteriol. 161:333–339.
17. Cregg, J. M., A. H. Nguyen, and J. Ito. 1980. DNA modification induced during infection of *Bacillus subtilis* by phage φ3T. Gene 12:17–24.
18. Cully, D. F., and Garro, A. J. 1980. Expression of superinfection immunity to bacteriophage φ105 by *Bacillus subtilis* cells carrying a plasmid chimera of pUB110 and *Eco*RI fragment F of φ105 DNA. J. Virol 34:789–791.
19. Cully, D. F., and A. J. Garro. 1985. Nucleotide sequence of the immunity region of *Bacillus subtilis* bacteriophage φ105: identification of the repressor gene and its mRNA and protein products. Gene 38:153–164.
20. Dean, D. H., M. Arnaus, and H. O. Halvorson. 1976. Genetic evidence that *Bacillus* bacteriophage φ105 integrates by insertion. J. Virol. 20:339–341.
21. Dean, D. H., C. L. Fort, and J. A. Hoch. 1978. Characterization of temperate phages of *Bacillus subtilis*. Curr. Microbiol. 1:213–217.
22. Dean, D. H., J. C. Orrego, K. W. Hutchison, and H. O. Halvorson. 1976. New temperate bacteriophage for *Bacillus subtilis*, ρ11. J. Virol. 20:509–519.
23. Derbyshire, V., and M. Belfort. 1998. Lightning strikes twice: intron–intein coincidence. Proc. Natl. Acad. Sci. USA 95:1356–1357.
24. Dhaese, P., M. R. Dobblelaere, and M. Van Montagu. 1985. The temperate *B. subtilis* phage φ105 genome contains at least two distinct regions encoding superinfection immunity. Mol. Gen. Genet. 200:490–492.
25. Dhaese, P., J. Seurinck, B. de Smet, and M. Van Montagu. 1985. Nucleotide sequence and mutational analysis of an immunity repressor gene from *Bacillus subtilis* temperate phage φ105. Nucleic Acids Res. 13:5441–5455.
26. East, A. K., and J. Errington. 1989. A new bacteriophage vector for cloning in *Bacillus subtilis* and the use of φ105 for protein synthesis in maxicells. Gene 81:35–43.
27. Edgell, D. R., M. Belfort, and D. A. Shub. 2000. Barriers to intron promiscuity in bacteria. J. Bacteriol. 182:5281–5289.
28. Ehrlich, S. D., I. Bursztyn-Pettegrew, I. Stroynowski, and J. Lederberg. 1976. Expression of the thymidylate synthetase gene of the *Bacillus subtilis* bacteriophage φ3T in *Escherichia coli.* Proc. Natl. Acad. Sci. USA 73:4145–4149.
29. Ellis, D. M., and D. H. Dean. 1985. Nucleotide sequence of the cohesive single-stranded ends of *Bacillus subtilis* temperate bacteriophage φ105. J. Virol. 55:513–515.
30. Errington, J. 1988. Generalized cloning vectors. Biotechnology 10:345–362.
31. Errington, J. 1993. Temperate phage vectors, pp. 645–650. *In* Sonenshein, A. L., J. A. Hoch, and R. Losick (eds.) *Bacillus subtilis* and Other Gram-positive Bacteria: Biochemistry, Physiology and Molecular Genetics. American Society for Microbiology, Washington, D.C.
32. Errington, J., and N. Pughe. 1987. Upper limit for DNA packaging by *Bacillus subtilis* bacteriophage φ105: isolation of phage deletion mutants by induction of oversized prophages. Mol. Gen. Genet. 210:347–351.
33. Estrela, A. I., H. de Lencastre, and L. J. Archer. 1986. Resistance of a *Bacillus subtilis* mutant to a group of temperate bacteriophages. J. Gen. Microbiol. 132:411–415.
34. Estrela, A. I., H. M. Pooley, H. de Lencastre, and D. Karamata. 1991. Genetic and biochemical characterization of *Bacillus subtilis* 168 mutants specifically blocked in the synthesis of the teichoic acid poly(3-*O*-β-D-glucopyranosyl-*N*-acetylgalactosamine 1-phosphate): *gneA*, a new locus, is associated with UDP-*N*-acetylglucosamine 4-epimerase activity. J. Gen. Microbiol. 137:943–951.
35. Fernandes, R. M., H. de Lencastre, and L. J. Archer. 1983. Two newly isolated temperate phages of *Bacillus subtilis*, Broteria-Genetics 4:27–33.
36. Fernandes, R. M., H. de Lencastre, and L. J. Archer. 1986. Three new temperate phages of *Bacillus subtilis*, J. Gen. Microbiol. 132:661–668.
37. Fernandes, R. M., H. de Lencastre, and L. J. Archer. 1989. Specialized transduction in *Bacillus subtilis* by the phages IG1, IG3 and IG4. Arch. Virol. 105:137–140.
38. Fernandes, R. M., H. de Lencastre, and L. J. Archer. 1990. Action of 6-(*p*-hydroxyphenylazo)-uracil on bacteriophage IG1. Arch. Virol. 113:177–181.
39. Fink, P. S., R. Z. Korman, J. M. Odebralski, and S. A. Zahler. 1981. *Bacillus subtilis* bacteriophage SPβ*cl* is a deletion mutant of SPβ. Mol. Gen. Genet. 182:514–515.
40. Fink, P. S., and S. A. Zahler. 1982. Specialized transduction of the *ilvD-thyB-ilvA* region mediated by *Bacillus subtilis* bacteriophage SPβ. J. Bacteriol. 150:1274–1279.
41. Fink, P. S., and S. A. Zahler. 1982. Restriction fragment maps of the genome of *Bacillus subtilis* bacteriophage SPβ. Gene 19:335–338.
42. Fink, P. S., and S. A. Zahler. 1983. SPβ *c2 pilvA*: plaque-forming bacteriophages that transduce the *Bacillus subtilis ilvA* gene. Abstracts of the 83rd Meeting of the American Society for Microbiology, p. 111.
43. Flock, J.-I. 1977. Deletion mutants of temperate *Bacillus subtilis* bacteriophage φ105. Mol. Gen. Genet. 155:241–247.
44. Foster, S. 1993. Analysis of *Bacillus subtilis* 168 prophage-associated lytic enzymes; identification and characterization of CWLA-related prophage proteins. J. Gen. Microbiol. 139:3177–3184.
45. Freeman, A. G., K. M. Schweikart, and L. L. Larcom. 1987. Effect of ultraviolet radiation on the *Bacillus subtilis* phages SPO2, SPP1 and φ29 and their DNAs. Mutat. Res. 184:187–196.
46. Gardner, A., J. Odebralski, S. Zahler, R. Z. Korman, and A. I. Aronson. 1982. Glutamine synthetase subunit mixing and regulation in *Bacillus subtilis* partial diploids. J. Bacteriol. 149:378–380.
47. Garro, A. J., H. Leffert, and J. Marmur. 1970. Genetic mapping of a defective bacteriophage on the chromosome of *Bacillus subtilis* 168. J. Virol. 6:340–343.

48. Ghim, S. Y., S. K. Choi, B. S. Shin, and S. H. Park. 1998. An 8 kb nucleotide sequence at the 3′ flanking region of the *sspC* gene (184°) on the *Bacillus subtilis* 168 chromosome containing an intein and an intron. DNA Res. 5:121–126.
49. Gibson, R. M., and J. Errington. 1992. A novel *Bacillus subtilis* expression vector based on bacteriophage φ105. Gene 121:137–142.
50. Graham, S., Y. Yoneda, and F. E. Young. 1979. Isolation and characterization of viable deletion mutants of *Bacillus subtilis* bacteriophage SP02. Gene 7:69–77.
51. Gunthert, U., R. Lauster, and L. Reiners. 1986. Multispecific DNA methyltransferases from *Bacillus subtilis* phages. Properties of wild-type and various mutant enzymes with altered DNA affinity. Eur. J. Biochem. 159:485–492.
52. Gunthert, U., and L. Reiners. 1987. *Bacillus subtilis* phage SPR codes for a DNA methyltransferase with triple sequence specificity. Nucleic Acids Res. 15:3689–3702.
53. Hartford, O. M., and B. C. A. Dowds. 1992. Cloning and characterization of genes induced by hydrogen peroxide in *Bacillus subtilis*. J. Gen. Microbiol. 138:2061–2068.
54. Hemphill, H. E., I. Gage, S. A. Zahler, and R. Z. Korman. 1980. Prophage-mediated production of a bacteriocinlike substance by SPβ lysogens of *Bacillus subtilis*. Can. J. Microbiol. 26:1328–1333.
55. Hendrix, R. W., M. C. M. Smith, R. N. Burns, M. E. Ford, and G. F. Hatfull. 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. Proc. Natl. Acad. Sci. USA 96:2192–2197.
56. Iismaa, T. P., C. M. Carrigan and R. G. Wake. 1988. Relocation of the replication terminus, *terC*, of *Bacillus subtilis* to a new chromosomal site. Gene 67:183–191.
57. Iismaa, T. P., and R. G. Wake. 1987. The normal replication terminus of the *Bacillus subtilis* chromosome, *terC*, is dispensable for vegetative growth and sporulation. J. Mol. Biol. 195:299–310.
58. Inselberg, J. W., T. Eremenko-Volpe, L. Greenwald, W. L. Meadow, and J. Marmur. 1969. Physical and genetic mapping of the SP02 prophage on the chromosome of *Bacillus subtilis*. J. Virol. 3:627–628.
59. Itaya, M. 1993. Stability and asymmetric replication of the *Bacillus subtilis* 168 chromosome structure. J. Bacteriol. 175:741–749.
60. Iwakura, M., M. Kawata, M. Tsuda, and T. Tanaka. 1988. Nucleotide sequence of the thymidylate synthetase B and dihydrofolate reductase genes contained in one *Bacillus subtilis* operon. Genetics. 64:9–20.
61. Karamata, D., H. M. Pooley, and M. Monod. 1987. Expression of heterologous genes for wall teichoic acid in *Bacillus subtilis* 168. Mol. Gen. Genet. 207:73–81.
62. Kawamura, F., H. Saito, and Y. Ikeda. 1979. Method for the construction of specialized transducing phage ρ11 of *Bacillus subtilis*. Gene 5:87–91.
63. Kawamura, F., H. Saito, Y. Ikeda, and J. Ito. 1979. Viable deletion mutants of *Bacillus subtilis* phage ρ11. J. Gen. Appl. Microbiol. 25:223–236.
64. Kenny, E., T. Atkinson, and B. S. Hartley. 1985. Nucleotide sequence of the thymidylate synthetase gene (*thyP3*) from the *Bacillus subtilis* phage φ3T. Gene 34:335–342.
65. Kobayashi, Y., and F. Kawamura. 1992. Molecular cloning. Biotechnology 22:123–141.
66. Krogh, S., S. T. Jorgensen, and K. M. Devine. 1998. Lysis genes of the *Bacillus subtilis* defective prophage PBSX. J. Bacteriol. 180:2110–2117.
67. Krogh, S., M. O'Reilly, N. Nolan, and K. M. Devine. 1996. The phage-like element PBSX and part of the *skin* element, which are resident at different locations on the *Bacillus subtilis* chromosome, are highly homologous. Microbiology 142:2031–2040.
68. Kroyer, J. M., J. B. Perkins, M. S. Rudinski, and D. H. Dean. 1980. Physical mapping of *Bacillus subtilis* phage ρ14 cloning vehicles: heteroduplex and restriction enzyme analyses. Mol. Gen. Genet. 177:511–517.
69. Kunst, F., N. Ogasawara, I. Moszer, et al. 1997. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. Nature 390:249–256.
70. Lange, C., A. Jugel, J. Walter, M. Noyer-Weidner, and T. A. Trautner. 1991. "Pseudo" domains in phage-encoded DNA methyltransferases. Nature 352:645–648.
71. Lange, C., M. Noyer-Weidner, T. A. Trautner, M. Weiner, and S. A. Zahler. 1991. M.H2I, a multispecific 5C-DNA methyltransferase encoded by *Bacillus amyloliquefaciens* phage H2. Gene 100:213–218.
72. Lazarevic, V. 1998. Is there a relic of a spliceosomal intron in *Bacillus subtilis* temperate phage SPβ? Mol. Microbiol. 29:1521–1528.
73. Lazarevic, V. 2001. Ribonucleotide reductase genes of *Bacillus* prophages: a refuge to introns and intein coding sequences. Nuceic Acids Res. 29:3212–3218.
74. Lazarevic, V., A. Düsterhöft, B. Soldo, H. Hilbert, C. Mauël, and D. Karamata. 1999. Nucleotide sequence of the *Bacillus subtilis* temperate bacteriophage SPβ*c*2. Microbiology 145:1055–1067.
75. Lazarevic, V., B. Soldo, A. Düsterhöft, B. Hilbert, C. Mauël, and D. Karamata. 1998. Introns and intein coding sequence in the ribonucleotide reductase genes of *Bacillus subtilis* temperate bacteriophage SPβ. Proc. Natl. Acad. Sci. USA 95:1692–1697.
76. Leung, Y. C., and J. Errington. 1995. Characterization of an insertion in the phage φ105 genome that blocks host *Bacillus subtilis* lysis and provides strong expression of heterologous genes. Gene 154:1–6.
77. Lipsky, R. H., R. Rosenthal, and S. A. Zahler. 1981. Defective specialized SPβ transducing bacteriophages of *Bacillus subtilis* that carry the *sup-3* or *sup-44* gene. J. Bacteriol. 148:1012–1015.
78. Longchamp, P. F., C. Mauël, and D. Karamata. 1994. Lytic enzymes associated with defective prophages of *Bacillus subtilis*: sequencing and characterization of the region comprising the *N*-acetylmuramoyl-L-alanine amidase gene of prophage PBSX. Microbiology 140:1855–1867.
79. Mackey, C. J., and S. A. Zahler. 1982. Insertion of bacteriophage SPβ into the *citF* gene of *Bacillus subtilis* and specialized transduction of the *ilvBC-leu* genes. J. Bacteriol. 151:1222–1229.
80. McLaughlin, J. R., H. C. Wong, Y. E. Ting, J. N. Van Arsdell, and S. Chang. 1986. Control of lysogeny and immunity of *Bacillus subtilis* temperate bacteriophage SPβ by its *d* gene. J. Bacteriol. 167:952–959.
81. McDonnell, G. E., and D. J. McConnell. 1994. Overproduction, isolation, and DNA-binding characteristics of Xre,

the repressor protein from the *Bacillus subtilis* defective prophage PBSX. J. Bacteriol. 176:5831–5834.

82. McDonnell, G. E., H. Wood, K. M. Devine, and D. J. McConnell. 1994. Genetic control of bacterial suicide: regulation of the induction of PBSX in *Bacillus subtilis*. J. Bacteriol. 176:5820–5830.
83. McVeigh, R. R., and R. E. Yasbin. 1996. Phenotypic differentiation of "smart" versus "naïve" bacteriophages of *Bacillus subtilis*. J. Bacteriol. 178:3399–3401.
84. Mele, J. 1972. Biological characterization and prophage mapping of a lysogenizing bacteriophage for *Bacillus subtilis*. PhD thesis, University of Massachusetts, Amherst, Mass.
85. Mizukami, T., F. Kawamura, H. Takahashi, and H. Saito. 1980. A physical map of the genome of the *Bacillus subtilis* temperate phage ρ11. Gene 11:157–162.
86. Moitoso de Vargas, L., and A. Landy. 1989. DNA looping generated by DNA bending protein IHF and the two domains of lambda integrase. Science 244:1457–1461.
87. Moszer, I., E. P. Rocha, and A. Danchin. 1999. Codon usage and lateral gene transfer in *Bacillus subtilis*. Curr. Opin. Microbiol. 2:524–528.
88. Murphy, F. A., C. M. Fauquet, D. H. L. Bishop, S. A. Ghabrial, A. W. Jarvis, G. P. Martelli, M. A. Mayo, and M. D. Summers (eds.) 1995. Virus taxonomy: classification and nomenclature of viruses. Sixth report of the International Committee on Taxonomy of Viruses. Arch. Virol. (Suppl) 10.
89. Nagy, E., B. Pragai, and G. Ivanovics. 1976. Characteristics of phage AP50, an RNA phage containing phospholipids. J. Gen. Virol. 32:129–132.
90. Neuhard, J., A. R. Price, L. Schack, and E. Thomassen. 1978. Two thymidylate synthetases in *Bacillus subtilis*. Proc. Natl. Acad. Sci. USA 75:1194–1198.
91. Noyer-Weidner, M., S. Jentsch, J. Kupsch, M. Bergbauer, and T. A. Trautner. 1985. DNA methyltransferase genes of *Bacillus subtilis* phages: structural relatedness and gene expression. Gene 35:143–150.
92. Noyer-Weidner, M., S. Jentsch, B. Pawlek, U. Günthert, and T. Trautner. 1983. Restriction and modification in *Bacillus subtilis*: DNA methylation potential of the related bacteriophages Z, SPR, SPβ, ϕ3T, and ρ11. J. Virol. 46:446–453.
93. Noyer-Weidner, M., and T. A. Trautner. 1992. DNA methylation: molecular biology and biological significance, pp. 39–108. *In* Jost, J. P., and H. P. Saluz (eds.) DNA Methylation: Molecular Biology and Significance. Birkhäuser Verlag, Basel.
94. Noyer-Weidner, M., J. Walter, P.-A. Terschüren, S. Chai, and T. A. Trautner. 1994. M.ϕ3TII: a new monospecific DNA (cytosine-C5) methyltransferase with pronounced amino acid sequence similarity to a family of adenine-N6-DNA methyltransferases. Nucleic Acids Res. 22:4066–4072.
95. Odebralski, J. M., and S. A. Zahler. 1982. Specialized transduction of the *kauA* and *citK* genes of *Bacillus subtilis* by bacteriophages ϕ3T. Abstracts of the 82nd Annual Meeting of the American Society for Microbiology, p. 130.
96. Okamoto, K., J. A. Mudd, and J. Marmur. 1968. Conversion of *B. subtilis* DNA to phage DNA following mitomycin C induction. J. Mol. Biol. 34:429–437.
97. Parker, A. P., and D. H. Dean. 1986. Temperate *Bacillus* bacteriophage SP16 genome is circularly permuted and terminally redundant. J. Bacteriol. 167:719–721.
98. Pečnková, T., and V. Pačes. 1999. Molecular phylogeny of ϕ29-like phages and their evolutionary relatedness to other protein-primed replicating phages and other phages hosted by Gram-positive bacteria. J. Mol. Evol. 48:197–208.
99. Petit, M.-A., and S. D. Ehrlich. 2000. The NAD-dependent ligase encoded by *yerG* is an essential gene of *Bacillus subtilis*. Nucleic Acids Res. 28:4642–4648.
100. Podvin, L., and M. Steinmetz. 1992. A *degU* containing SPβ prophage complements superactivator mutations affecting the *Bacillus subtilis degSU* operon. Res. Microbiol. 143:559–567.
101. Poth, H., and P. Youngman. 1988. A new cloning system for *Bacillus subtilis* comprising elements of phage, plasmid and transposon vectors. Gene 73:215–226.
102. Regamey, A., and D. Karamata. 1998. The *N*-acetylmuramoyl-L-alaline amidase encoded by the *Bacillus subtilis* 168 prophage SPβ. Microbiology 144:885–893.
103. Regamey, A., V. Lazarevic, P. Hauser, and D. Karamata. 2000. Study of chromosome rearrangements associated with the trpE26 mutation of *Bacillus subtilis*. Mol. Mircobiol. 36:1234–1249.
104. Robertson, C. A., and H. A. Nash. 1988. Bending of the bacteriophage lambda attachment site by *Escherichia coli* integration host factor. J. Biol. Chem. 263:3554–3557.
105. Romig, W. R. 1968. Infectivity of *Bacillus subtilis* bacteriophage deoxyribonucleic acids extracted from mature particles and from lysogenic hosts. Bacteriol. Rev. 32:349–357.
106. Rosenthal, R., P. A. Toye, R. Z. Korman, and S. A. Zahler. 1979. The prophage of SPβ*c*2d*citK*$_1$, a defective specialized transducing phage of *Bacillus subtilis*. Genetics 92:721–739.
107. Ross, W., and A. Landy. 1983. Patterns of λ int recognition in the regions of strand exchange. Cell 33:261–272.
108. Rutberg, L. 1969. Mapping of a temperate bacteriophage active on *Bacillus subtilis*. J. Virol. 3:38–44.
109. Rutberg, L., B. Raden, and J.-I. Flock. 1981. Cloning and expression of bacteriophage SPO2 DNA polymerase gene L in *Bacillus subtilis*, using the *Staphylococcus aureus* plasmid pC194. J. Virol. 39:407–412.
110. Santos, I., and H. de Lencastre. 1992. Cloning of the thymidylate synthetase gene (*thyPIG3*) from the *Bacillus subtilis* temperate phage IG3. Arch. Virol. 127:65–74.
111. Scher, B. M., D. H. Dean, and A. J. Garro. 1977. Fragmentation of *Bacillus* bacteriophage ϕ105 DNA by restriction endonuclease *Eco*RI: evidence for complementary single-stranded DNA in the cohesive ends of the molecule. J. Virol. 23:377–383.
112. Sciochetti, S. A., P. J. Piggot, and G. W. Blakely. 2001. Identification and characterization of the *dif* site from *Bacillus subtilis*. J. Bacteriol. 183:1058–1068.
113. Schmeiger, H., K. M. Taleghani, A. Meierl, and L. Weiss. 1990. A molecular analysis of terminase cuts in headful packaging of *Salmonella* phage P22. Mol. Gen. Genet. 221:199–202.

114. Shapiro, J. M., D. H. Dean, and H. O. Halvorson. 1974. Low-frequency specialized transduction with *Bacillus subtilis* bacteriophage φ105. Virology 62:393–403.
115. Smith, I., and H. Smith. 1973. Location of the SP02 attachment site and the bryamycin resistance marker on the *Bacillus subtilis* chromosome. J. Bacteriol. 114:1138–1142.
116. Soldo, B., V. Lazarevic, P. Margot, and D. Karamata. 1993. Sequencing and analysis of the divergon comprising *gtaB*, the structural gene of UDP-glucose pyrophosphorylase of *Bacillus subtilis* 168. J. Gen. Microbiol. 139:3185–3195.
117. Spancake, G. A., S. D. Daignault, and H. E. Hemphill. 1987. Gene homology and divergence in the SPβ-related bacteriophages of *Bacillus subtilis*. Can. J. Microbiol. 33:249–255.
118. Spancake, G. A., and H. E. Hemphill. 1985. Deletion mutants of *Bacillus subtilis* bacteriophage SPβ. J. Virol. 55:39–44.
119. Spancake, G. A., H. E. Hemphill, and P. S. Fink. 1984. Genome organization of SPβ c2 bacteriophage carrying the *thyP* gene. J. Bacteriol. 157:428–434.
120. Steensma, H. Y. 1981. Effect of defective phages on the cell membrane of *Bacillus subtilis* and partial characterization of the phage protein involved in killing. J. Gen. Virol. 56:275–286.
121. Steensma, H. Y., L. A. Robertson, and J. D. van Elsas. 1978. The occurrence and taxonomic value of PBS-X-like defective phages in the genus *Bacillus.* Antonie van Leeuwenhoek 44:353–366.
122. Stickler, D. J., R. G. Tucker, and D. Kay. 1965. Bacteriophage-like particles released from *Bacillus subtilis* after induction with hydrogen peroxide. Virology 26: 142–145.
123. Stroynowski, I. 1981. Integration of the bacteriophage φ3T-coded thymidylate synthetase gene into the *Bacillus subtilis* chromosome. J. Bacteriol. 148:101–108.
124. Takemaru, K.-I., M. Mizuno, T. Sato, M. Takeuchi, and Y. Kobayashi. 1995. Complete nucleotide sequence of a *skin* element excised by DNA rearrangement during sporulation in *Bacillus subtilis.* Microbiology 141:323–327.
125. Tam, N. H. and R. Borriss. 1995. The *thyA* gene from *Bacillus subtilis* exhibits similarity with the phage φ3T thymidylate synthetase gene. Microbiology 141:291–297.
126. Terschüren, P. A., M. Noyer-Weidner, and T. A. Trautner. 1987. Recombinant derivatives of *Bacillus subtilis* phage Z containing the DNA methyltransferase genes of related methylation proficient phages. J. Gen. Microbiol. 133:945–952.
127. Thornewell, S. J., A. K. East, and J. Errington. 1993. An efficient expression and secretion system based on *Bacillus subtilis* phage φ105 and its use for the production of *B. cereus* β-lactamase I. Gene 133:47–53.
128. Tran-Betcke, A., B. Behrena, M. Noyer-Weidner, and T. A. Trautner. 1986. DNA methyltransferase genes of *Bacillus subtilis* phages: comparison of their nucleotide sequences. Gene 42:89–96.
129. Trautner, T. A., T. Balganesh, K. Wilke, M. Noyer-Weidner, E. Rauhut, R. Lauster, B. Behrens, and B. Pawlek. 1988. Organization of target-recognizing domains in the multispecific DNA (cytosine-5) methyltransferases of *Bacillus subtilis* phages SPR and φ3T. Gene 74:267.
130. Trautner, T. A., B. Pawlek, U. Gunthert, U. Canosi, S. Jentsch, and M. Freund. 1980. Restriction and modification in *Bacillus subtilis*: identification of a gene in the temperate phage SPβ coding for a *BsuR* specific modification methyltransferase. Mol. Gen. Genet. 180:361–367.
131. Tsutsumi, Y., H. Hirokawa, and K. Shishido. 1990. A new defective phage containing a randomly selected 8 kilobase-pairs fragment of host chromosomal DNA inducible in a strain of *Bacillus natto.* FEMS Microbiol. Lett. 60:41–46.
132. Tucker, R. G. 1969. Acquisition of thymidylate synthetase activity by a thymine-requiring mutant of *Bacillus subtilis* following infection by the temperate phage φ3T. J. Gen. Virol. 4:489–504.
133. Van Kaer, L., Y. Gansemans, M. Van Montagu, and P. Dhaese. 1988. Interaction of the *Bacillus subtilis* phage φ105 repressor with operator DNA: a genetic analysis. EMBO J. 7:859–866.
134. Van Kaer, L., M. Van Montagu, and P. Dhaese. 1987. Transcriptional control in the *Eco*RI-F immunity region of *Bacillus subtilis* phage φ105: identification and unusual structure of the operator. J. Mol. Bio. 197:55–67.
135. Van Kaer, L., M. Van Montagu, and P. Dhaese. 1989. Purification and in vitro DNA-binding specificity of the *Bacillus subtilis* phage φ105 repressor, J. Biol. Chem. 264:14784–14791.
136. Varon, D., S. A. Boylan, K. Okamoto, and C. W. Price. 1993. *Bacillus subtilis gtaB* encodes UDP-glucose pyrophosphorylase and is controlled by stationary-phase transcription factor σ^B. J. Bacteriol. 175:3964–3971.
137. Warner, F. D., G. A. Kitos, M. P. Romano, and H. E. Hemphill. 1977. Characterization of SPβ: a temperate bacteriophage from *Bacillus subtilis* 168M. Can. J. Microbiol. 23:45–51.
138. Weiner, M. P. 1986. Characterization of phage H2, PhD thesis, Cornell University, Ithaca, N.Y.
139. Weiner, M. P., and S. A. Zahler. 1988. Genome homology and host range of some SPβ-related bacteriophages of *Bacillus subtilis* and *Bacillus amyloliquefaciens*. J. Gen. Virol. 69:1307–1316.
140. Weisberg, R. A., and A. Landy. 1983. Site-specific recombination in phage lambda, pp. 211–250. *In* R. W. Hendrix, J. W. Roberts, F. W. Stahl, and R. A. Weisberg (eds.) *Lambda II*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
141. Wilke, K., E. Rauhut, M. Noyer-Weidner, R. Lauster, B. Pawlek, B. Behrens, and T. Trautner. 1988. Sequential order of target-recognizing domains in multispecific DNA-methyltransferases. EMBO J. 7:2601-2609.
142. Williams, M. T., and F. E. Young. 1977. Temperate *Bacillus subtilis* bacteriophage φ3T: chromosomal attachment site and comparison with temperate bacteriophages φ105 and SP02. J. Virol. 21:522–529.
143. Wilson, G. A., M. T. Williams, H. W. Baney, and F. E. Young. 1974. Characterization of temperate bacteriophages of *Bacillus subtilis* by the restriction endonuclease *Eco*RI: evidence for three different temperate bacteriophages. J. Virol. 14:1013–1016.
144. Wood, H. E., M. T. Dawson, K. M. Devine, and D. J. McConnell. 1990. Characterization of PBSX, a defective prophage of *Bacillus subtilis*. J. Bacteriol. 172:2667–2674.

145. Wood, H. E., K. M. Devine, and D. J. McConnell. 1990. Characterization of a repressor gene (*xre*) and a temperature-sensitive allele from the *B. subtilis* prophage, PBSX. Gene 96:83–88.
146. Yang, Q. 1989. Characterization of the temperate *Bacillus subtilis* bacteriophage SPβ attachment site. MS thesis, Wright State University, Dayton, OH.
147. Yasbin, R. E., D. Cheo, and K. W. Bayles. 1991. The SOB system of *Bacillus subtilis*: a global regulon involved in DNA repair and differentiation. Res. Microbiol. 142:885–892.
148. Yasbin, R. E., P. I. Fields, and B. J. Andersen. 1980. Properties of *Bacillus subtilis* 168 derivatives freed of their natural prophages. Gene 12:155–159.
149. Young, M., C. Mauel, P. Margot, and D. Karamata. 1989. Pseudo-allelic relationship between non-homologous genes concerned with biosynthesis of polyglycerol phosphate and plyribitol phosphate teichoic acids in *Bacillus subtilis* strains 168 and W23. Mol. Microbiol. 3:1805–1812.
150. Zahler, S. A. 1982. Specialized transduction in *Bacillus subtilis*, pp. 269–305. *In* D. A. Dubnau (ed.) *Molecular Biology of the Bacilli*, vol. 1. Academic Press, New York.
151. Zahler, S. A. Temperate bacteriophages of *Bacillus subtilis*, pp. 559–592. *In* R. Calendar (ed.) *The Bacteriophages*, vol. 1 Plenum Press, New York.
152. Zahler, S. A. 1993. Temperate bacteriophages, pp. 831–842. *In* A. Sonenshein, J. Hoch and R. Losick (eds.) *Bacillus subtilis and Other Gram-positive Bacteria: Biochemistry, Physiology and Molecular Genetics*. American Society for Microbiology, Washington, D.C.
153. Zahler, S. A., R. Z. Korman, C. Thomas, P. S. Fink, M. P. Weiner, and J. M. Odebralski. 1987. H2, a temperate bacteriophage isolated from *Bacillus amyloliquefaciens* strain H. J. Gen. Microbiol. 133:2937–2944.
154. Zahler, S. A., R. Z. Korman, R. Rosenthal, and H. E. Hemphill. 1977. *Bacillus subtilis* bacteriophage SPβ: localization of the prophage attachment site, and specialized transduction. J. Bacteriol. 129:556–558.
155. Zahler, S. A., R. Z. Korman, C. Thomas, and J. M. Odebralski. 1987. Temperate bacteriophages of *Bacillus amyloliquefaciens*. J. Gen. Microbiol. 133:2933–2935.
156. Zuber, P. 2001. A peptide profile of the *Bacillus subtilis* genome. Peptides 22:1555–1557.

36

Phages of *Lactococcus lactis*

LONE BRØNDSTED
KARIN HAMMER

Lactococci are Gram-positive mesophilic bacteria with low G-C content that belong to the group of lactic acid bacteria. They are aerotolerant and live by means of fermentation, as they are lacking a respiratory chain, and the main end product during fermentation is lactic acid. In nature lactococci are mainly found on plant material (100). Some lactococci have apparently adapted to multiply in milk and contain a plasmid-encoded milk protease and enzymes responsible for catabolism of lactose. *Lactococcus lactis* is used world-wide as starter culture for large-scale milk fermentations producing cheese such as cheddar and other fermented milk products. It has been estimated that approximately 10^7 tonnes of cheese are made annually, leading to human consumption of close to 10^{18} lactococci (41). In the genus *Lactococcus* five species are currently found: *L. lactis, L. garviae, L. plantarum, L. piscium, L. raffinolactis* (101, 118). Of these only *L. lactis* is used as starter culture and only phages from *L. lactis* have been studied at the molecular level. *L. lactis* is divided into three subspecies. Of these only *L. lactis* ssp. *lactis* and *L. lactis* ssp *cremoris* are used for milk fermentations. They are distinguished by differences in their DNA sequences, including those encoding 16S rRNA (45).

The phylogenetic trees based on 16S rRNA identify streptococci as the closest relatives to *L. lactis*. In contrast to the many streptococcal species, lactococci are nonpathogenic, food-grade bacteria and may even be beneficial to health. Their potential for new applications such as oral vaccines is being investigated (106). In the last decade an impressive amount of research has been conducted on *L. lactis* and it is now considered a model organism. As laboratory model strains, IL1403 (*L. lactis* ssp. *lactis*) and MG1363 (*L. lactis* ssp. *cremoris*) are widely used. The 2.4 Mbp genome of IL1403 has been sequenced (12) and genome sequencing of MG1363 is in progress (62). Genetic techniques and tools such as transduction, conjugation, transformation, and transposon mutagenesis are available (39). Natural competence has never been described for lactococci despite all the necessary competence genes having been found on the IL1403 chromosome (12).

The industrial use of *L. lactis* in vast amounts for milk fermentations is providing a gigantic large-scale environment for phage reproduction and evolution. Phage infections are difficult to avoid, since pasteurized milk is not sterile and may contain phages that potentially destroys the fermentation and hence the product. Furthermore, the use of defined strains as starter cultures has greatly limited the number of industrial strains used and hence made it easier for phage contaminants to proliferate. After the dairy production failures in the mid-1930s were recognized to be caused by phage infections (117), research into phages and phage resistance has been a major issue in the lactococcal field. Phage resistance systems, however, are outside the scope of this chapter; for reviews see (2, 40).

In 1949 it was discovered that *L. lactis* could be lysogenic (98). Later, lysogeny was found to be a very common phenomenon in *L. lactis* strains (47, 51, 53). In accordance with this, several prophages were discovered when the genome of the model strain *L. lactis* ssp. *lactis* IL1403 was sequenced (30). The majority of phages infecting *L. lactis* belong to the *Siphoviridae* family and carry a long, noncontractile tail, while a few belongs to the *Podoviridae* family and have a very short tail. Among the lactococcal phages only morphotypes B1, B2, C2, and C3 have been observed (table 36-1; see also chapter 2 for a general discussion of phage classification). At present only phages containing double-stranded DNA genomes have been identified. On the basis of morphology and DNA hybridizations (55) lactococcal phages were divided into 12 phage species. Among these, phage species 1483, 1358, and øT187 have later been assigned to the P335 phage species (30, 54, 83). We have furthermore included phage BK5-T in the P335 species (30). The remaining eight species, their type phage, and some examples of specific phage members are shown in table 36-1. Over the years many lactococcal phages have been isolated world-wide and more than 80% of all lactococcal phages

Table 36-1 Taxonomy of Lactococcal Phages as Modified from (55)

Family	Morphotype[a]	Phage species	Type phage	Members
Siphoviridae	B1	936	P008	P008, F4-1, sk1, ϕUS3, bIL41, bIL66, bIL170, uc1001, uc1002, 712
	B1	P335[b]	P335	Φ31, Φ50, ul36, r1t, ΦLC3, TP901-1, Tuc2009, C3-T1, Q30, Q33, 7-9, 1483, 1358, ϕT187, bIL285, bIL286, bIL309, TPW22, 4268, BK5-T[c]
	B1	P107	P107	
	B1	P087	P087	
	B1	949	949	
	B2	c2	c6A	c2, c6A, bIL67, bIL188, ϕvML3, P001, P6, ϕ197
Podoviridae	C2	P034	P034	P034, ascc ϕ28 (I. B. Powell, personal comunication)
	C3	KSY1	KSY1	

[a] For a description of morphotypes see chapter 2.

[b] For further subdivision of the P335 species see table 36-5.

[c] BK5-T was originally proposed to be a separate phage species.

Table 36-2 Lactococcal Phages with Fully Sequenced Genomes

Phage species	Phage	Type	Ends	Genome size (bp)	Reference	Accession no.
c2	bIL67	Virulent	*cos*	22,195	(102)	L33769
	c2	Virulent	*cos*	22,172 or 22,163	(71)	L48605
936	sk1	Virulent	*cos*	28,451	(27)	AF011378
	bIL170	Virulent	*cos*	31,754	(34)	AF009630
P335	r1t	Temperate	*cos*	35,550	(111)	U38906
	TP901-1	Temperate	*pac*	37,667	(23)	AF304433
	Tuc2009	Temperate	*pac*	38,347	(36)	AF109874
	bIL285	Temperate		35,538	(30)	AF323668
	bIL286	Temperate		41,834	(30)	AF323669
	bIL309	Temperate		36,949	(30)	AF323670
	BK5-T	Temperate	*cos*	40,003	(35, 78)	AF176025
	ul36	Virulent		36,798	(65)	AF349457
	4268	Virulent		36,596		AF489521
Unknown	bIL310	Temperate		14,957	(30)	AF323671
	bIL311	Temperate		14,510	(30)	AF323672
	bIL312	Temperate		15,179	(30)	AF323673
P034	Asccϕ28	Virulent		18,762	I. B. Powell, personal communication	

isolated from the dairy environment belong to phage species 936, P335, and c2 (52, 60, 97). Due to their industrial importance, lactococcal phage research has been focused on these three phage species, and 16 of the 17 fully sequenced genomes belong to these species (table 36-2). In this chapter we will therefore concentrate on phages belonging to the c2, 936, and P335 phage species with the main emphasis on the sequenced members of these phages.

Phage Species c2

General Description

This group of phages consists of prolate-headed phages with long, noncontractile tails. So far only virulent *cos* phages with genome sizes from 18 to 22 kb have been assigned to this phage species. Phage bIL67 was the first lactococcal phage to be sequenced (102) and it was later found that bIL67 and phage c2 are highly related, since the two phage genomes share 80% nucleotide sequence identity (71). Genome sequencing of the c2 phage identified 39 open reading frames (ORFs) organized in two clusters of divergent orientations (figure 36-1) (71).

Transcription

Transcriptional analysis of the phage c2 genome during one-step growth showed two classes of genes: early and late (figure 36-1). Early mRNAs were found 5 and 10 minutes after infection as well as later in infection, while late mRNAs were found from 15 minutes onwards until cell lysis at 45 minutes (5). This temporal gene expression corresponds

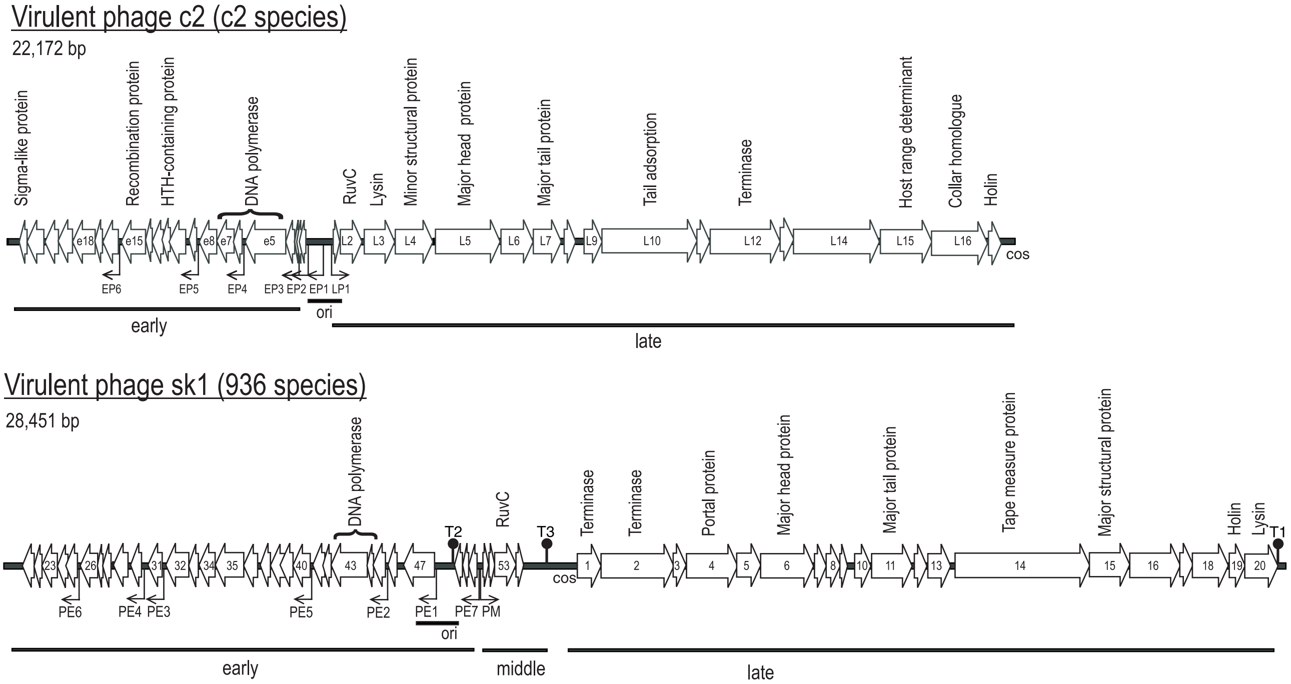

Virulent phage c2 (c2 species)
22,172 bp
Sigma-like protein
Recombination protein
HTH-containing protein
DNA polymerase
RuvC
Lysin
Minor structural protein
Major head protein
Major tail protein
Tail adsorption
Terminase
Host range determinant
Collar homologue
Holin
e18
e15
e8
e7
e5
L2
L3
L4
L5
L6
L7
L9
L10
L12
L14
L15
L16
cos
EP6
EP5
EP4
EP3
EP2
EP1
LP1
ori
early
late
Virulent phage sk1 (936 species)
28,451 bp
DNA polymerase
RuvC
Terminase
Terminase
Portal protein
Major head protein
Major tail protein
Tape measure protein
Major structural protein
Holin
Lysin
T2
T3
T1
23
26
31
32
34
35
40
43
47
53
cos
1
2
3
4
5
6
8
10
11
13
14
15
16
18
19
20
PE6
PE4
PE3
PE5
PE2
PE1
PE7
PM
ori
early
middle
late

Figure 36-1 Genome organization of virulent phages c2 and sk1. The DNA sequence of c2 is obtained from (71), accession number L48605, while the DNA sequence of phage sk1 is obtained from (27), accession number AF011378. Arrows indicate the size and direction of transcription of open reading frames (ORFs). Assigned functions are written above the ORFs. Identified and putative promoters are indicated as small black arrows, while origins of replication are shown as a black box below the ORFs. The locations of *cos* sites are indicated. Regions of the genome transcribed early, middle, or late in the infection cycle are marked with black lines below the genome.

to the two gene clusters found by sequence analysis (71). Six early and one late promoter were also suggested by sequence analysis, and the mRNA ends were verified by primer extension analysis according to unpublished results from the Jarvis group.

Gene Functions and Origin of Replication

Amino acid similarity suggests that phage c2 encodes a DNA polymerase by three successive ORFs (e5 to e7), whereas regulatory proteins, e12 and e22, were proposed due to the presence of a DNA-binding or a sigma-like motif, respectively. Finally, the e15 protein shows similarity to the recombination protein, Erf, of the *Salmonella* phage P22 (figure 36-1) (the biology of phage P22 is reviewed in chapter 29). Generally the ORFs identified in the early region were small in size and the possibility for creation of larger reading frames through the mechanisms of frame shifting has been discussed (71).

The identity of several of the structural genes in the late gene cluster was accomplished by a combination of SDS-PAGE, N-terminal sequencing, western blot, and immunogold electron microscopy of c2 phage particles (71). Gene products L5, L7, and L10 were thus identified as the major head, major tail, and putative tail adsorption protein, respectively. Analysis of several major virion protein bands with identical N-termini showed that the major head protein of c2 most likely is both processed proteolytically and covalently linked in protein complexes with molecular weights corresponding mainly to trimeric and hexameric forms of the processed major head subunit (71). Furthermore, a proteolytic cleavage site of the major head protein was predicted by computer analysis (35). A minor processed form of the major tail protein (L7) was also observed. Terminase function was assigned to L12 on the basis of domain homology (71). L3 had previously been identified as the phage lysin (115). A possible holin function was assigned to L17 on the basis of possession of two predicted transmembrane domains and a charged C-terminus (see chapter 10 for a review of holin structure and function), while L2 located immediately upstream of the lysin gene was suggested to encode a structure-dependent endonuclease activity necessary for maturation and packaging of phage DNA, as shown for phage bIL66 (7). A 521 bp fragment of the noncoding region located between the divergently transcribed gene clusters was shown to act as an origin of DNA replication in *L. lactis* when cloned in an origin screening vector, suggesting that this region harbors the origin of phage c2 replication (116). Investigation of replicating intermediates of c2 confirmed this, the data furthermore revealed a theta replication mechanism (24).

Relationship Between c2 and bIL67

Analysis of the DNA sequences of c2 and bIL67 showed that the sequences are almost identical (35, 71). Only three regions showed nonalignment: one corresponds to the origin of replication, another was located in a central domain in the putative tail adsorption gene, and the third covered a region encoding minor structural proteins (L14, L15, and L16 in c2) that, despite their nucleotide-sequence divergence, retained significant amino acid similarity to the corresponding proteins from phage bIL67. The L15 and L16 proteins were found to share significant amino acid similarity with putative tail adsorption proteins from both *cos*- and *pac*-site *Streptococcus thermophilus* phages. Experimental evidence has now been provided showing that the host range of both bIL67 and also another prolate phage is determined by protein 35, corresponding to the L15 protein of c2 (108) (figure 36-1). L16 has been identified as part of the collar structure of a c2 derivative possessing a collar. The collar-containing c2 phage harbors an additional gene (*col*) that is downstream of *L16* and the presence of both *col* and *L16* seems to be necessary for collar formation. Collarless c2 phages arise spontaneously from the collared form, presumably by homologous recombination between the *L16* and *col* genes (I. B. Powell, personal communication).

Phage Receptors

The only lactococcal phage receptor characterized at the molecular level is the membrane protein Pip (phage infection protein) that was required for infection by prolate-headed phages of the c2 species (44). It was suggested that phage absorption is a two-stage process involving a reversible adsorption to carbohydrate components in the cell wall followed by an irreversible interaction with Pip (84). Lactococcal host cells containing a deletion of the *pip* gene were found to be resistant to phages of the c2 species. Furthermore, no growth defects were observed (42). In contrast, Pip was found not to be required for infection by phages of the 936, P335, and 949 species (63). The corresponding phage protein (the antireceptor) involved in interaction with the Pip protein was identified for two of the prolate-headed c2 phages as mentioned above (108).

936 Phage Species

General Description

The 936 phage species contains small isometric-headed phages with short tails. Genome sizes range from 25 to 40 kb and at present only virulent *cos* phages have been identified. So far only one type of *cos* site has been identified (11 bp 3′ overhang CACAAAGGACT) (28, 91, 95, 99). The major parts of sequenced regions of 936 phages have been found to be highly similar and two members of the 936 species have been fully sequenced: phages sk1 (27) and bIL170 (34). As described in (35), their genomes could be aligned essentially over their entire lengths. The exceptions were mainly insertions in either one or the other phage. Since more functional studies have been performed on sk1, this phage will be used as model phage for the 936 phages. The genome of sk1 is organized in three gene clusters (figure 36-1). The two divergently located gene clusters mainly contain small genes of unknown function, while the remaining 20 genes encode proteins involved in DNA packaging, morphogenesis, and cell lysis (27).

Transcription

Transcriptional analysis of sk1 divided the gene expression during the lytic cycle into early, middle and late transcripts; the early transcripts were repressed in the later phases of the lytic cycle (26). The early region covers the large gene cluster consisting of 30 small ORFs, the middle region covers four small ORFs, and the late region consists of the 20 ORFs located downstream of the *cos* site (figure 36-1) (27).

In the early region seven partially overlapping transcripts were found, and the 5′ end of three of these (E1, E5, and E7) was mapped by primer extensions. Promoter activity of PE1 and PE7 was furthermore verified by promoter cloning. In addition, site-directed mutagenesis was shown to destroy promoter function of PE1 and PE7 (27). Sequences showing some similarity to lactococcal consensus promoters could be found in four additional intergenic regions in the early gene cluster. Promoter activity, however, was not demonstrated.

The cloned early promoters were active in the absence of phage proteins; in contrast a middle promoter (P_M) was identified, which was only functional in the presence of phage proteins. The DNA region necessary for P_M activity was further defined by deletion analysis and mutagenesis studies. The deletion analysis demonstrated that the region located at −46 to −55 upstream of the transcriptional start site was necessary for promoter activity and mutagenesis studies expanded the region necessary for promoter activity to cover from −36 to −55 (27). An identical P_M promoter had earlier been identified in phage bIL66 (6).

Northern analysis during sk1 infection showed the presence of seven early mRNAs, nine small middle mRNAs, and four large late mRNAs. The largest of the late transcripts, estimated in the range of 14 kb, may cover all the late genes, while the remaining late mRNA species are overlapping and probably are processed forms of the large primary transcript. A large 16 kb transcript from the late promoter in bIL41 has been demonstrated to cover all the late genes. The partially overlapping smaller late transcripts were suggested to be products from partial termination or from processing of the larger transcript. Possible RNase E sites were found at the end of some of these mRNAs (91).

Gene Functions and Origin of Replication

In the early region of the sk1 genome many of the small ORFs are overlapping and the opportunity for larger ORFs by programmed frame shifting of translation was suggested for some of these (27), as has been observed for other phages, such as for λ (67). Phage sk1 ORF43 and ORF44 proteins showed similarity to a protein found in the prolate-headed lactococcal phages: the proposed DNA polymerase subunit encoded by *e5* from phage c2 and *orf3* from phage bIL67. Although the genes flanking the c2 and bIL67 homologs of sk1 gene 43/44 are also annotated as DNA polymerase subunit genes, no matches to these flanking DNAP genes of c2 and bIL67 are found in sk1.

Sequence analysis of the encoded proteins in sk1 have identified the small terminase subunit, the large terminase subunit, portal protein, major head protein, major tail protein, tape measure protein, holin, and lysin to ORFs 1, 2, 4, 6, 11, 14, 19, and 20, respectively (figure 36-1) (19, 27). It was noted that the gene order in the structural genes of sk1 and λ was highly conserved except for the inverted order of ORF9 and ORF10, as well as for ORF17 and ORF18. In addition, no homologs of λ genes *D*, *J*, and *Rz* could be found (27). See chapter 27 for a review of these phage λ genes and gene order.

The function of the holin and lysin from sk1 was shown by the damage caused by induction of these ORFs in *Escherichia coli* (27). The major tail protein homolog had previously been demonstrated to be an abundant structural protein of the *L. lactis* ssp. *cremoris* phage F4-1 (33). In contrast to the c2 phage species, no evidence has been found for proteolysis of the major head protein in sk1 and other 936 phages, either experimentally or from analysis of the amino acid sequence (35).

The origin of replication was localized in the early region to an 800 bp fragment of sk1, which enabled plasmid replication in *L. lactis* (figure 36-1). This minimal origin of replication covered the intergenic region between *orf47* and *orf48*, including the PE1 promoter and the N-terminal 179 amino acids of protein ORF47, which has a total length of 231 amino acids. Three pairs of direct repeats and a pair of 67 bp repeated sequences were identified in the intergenic region (27). Currently it has not been determined whether it is the protein part of ORF47 versus DNA sequences within *orf47* that are required for replication.

Interaction with Abortive Infection Mechanisms

A region of phage bIL66, showing high similarity to sk1 (85% identity), was shown to be a target for the AbiD1 abortive infection locus. The region is expressed from the P_M promoter and contains ORF1 through ORF4 corresponding to ORF51 through ORF54, respectively, of sk1 (figure 36-1). Mutants of bIL66 overcoming the inhibition by AbiD1 were shown to be located in *orf1*. ORF2 and ORF3 of bIL66 have been shown to cause double-stranded DNA breaks in branched DNA structures. ORF3 furthermore has amino acid similarity to the *E. coli* RuvC Holliday junction resolvase. It was therefore suggested that ORF2 and ORF3 form a structure-dependent endonuclease necessary for maturation and packaging of the phage DNA (7). The current hypothesis is that AbiD1 acts in conjunction with ORF1 to reduce the amount of ORF3 below that required for normal phage development (7). It was proposed that translation was reduced, since transcription was not affected (6). The RuvC homolog was further identified in the prolate phages c2 (L2) and bIL67 (ORF23). Interestingly, AbiD1 also inhibits proliferation of the prolate-headed phages (3).

A transcriptional study during the lytic cycle of bIL170 confirmed the findings from sk1 of early, middle, and late expressed regions (90). The sensitivity of this phage to the *abiB* abortive infection locus was furthermore examined. Transcriptional analysis showed that when AbiB-containing host cells where infected with bIL170, an RNase was produced, destroying phage mRNA. The results suggest that an early-expressed phage protein in combination with AbiB forms or activates this RNase (90). It may be noted that phage bIL41, in contrast to phage bIL170, is resistant to the *abiB* abortive infection locus.

Difference in sensitivity against an Abi system has also been found for the 936 phages, sk1 and 712. A 324 bp region covering the *cos*-site from the resistant sk1 phage cloned in a multicopy plasmid was shown to protect the sensitive 712 phage from the *abiF* abortive infection locus (99).

P335 Phage Species

General Description

Phages of the P335 phage species are small isometric-headed phages with genomes ranging from approximately 30 to 42 kb. In this group of phages *cos* or *pac* sites are used during initiation of DNA packaging and as the only lactococcal phage species it contains both virulent and temperate phages. Nine genomes, including BK5-T, of both temperate and virulent members are completely sequenced, including three full-size prophages located in the *L. lactis* ssp. *lactis* IL1403 genome (table 36-2) (30). In addition, three satellite prophages from IL1403 (table 36-2) and parts of many other phage genomes are available from the databases. The genetic organization of the P335 phage genomes is highly similar (18, 23, 35, 65, 112). The temperate phages all have a small lysogenic operon and a large,

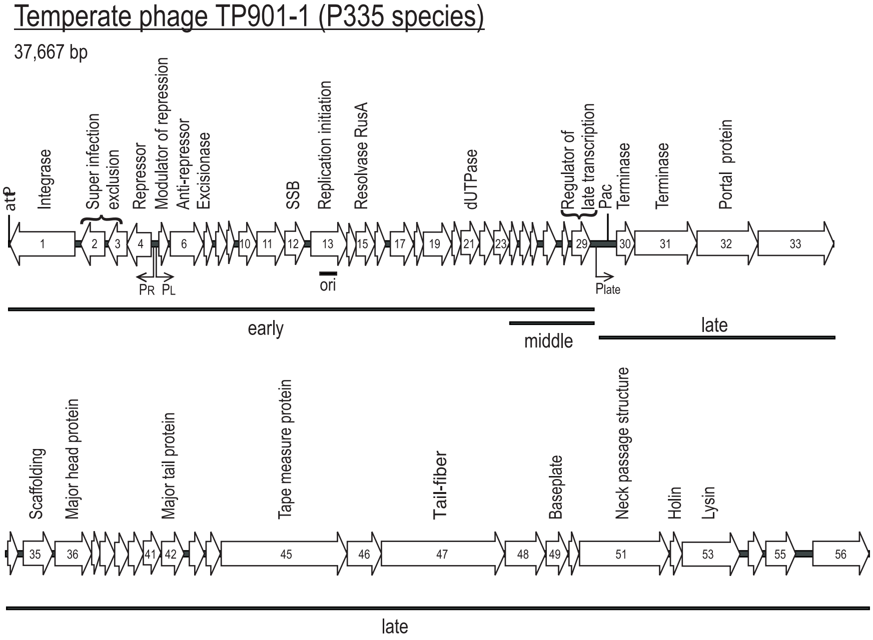

Temperate phage TP901-1 (P335 species)
37,667 bp
attP
Integrase
Super infection exclusion
Repressor
Modulator of repression
Anti-repressor
Excisionase
SSB
Replication initiation
Resolvase RusA
dUTPase
Regulator of late transcription
Pac
Terminase
Terminase
Portal protein
1
2
3
4
6
10
11
12
13
15
17
19
21
23
29
30
31
32
33
PR
PL
ori
Plate
early
middle
late
Scaffolding
Major head protein
Major tail protein
Tape measure protein
Tail-fiber
Baseplate
Neck passage structure
Holin
Lysin
35
36
41
42
45
46
47
48
49
51
53
55
56
late

Figure 36-2 Genome organization of temperate phage TP901-1. Sequence is obtained from (23), accession number AF304433. Arrows indicate the size and direction of transcription of open reading frames. Assigned functions are written above the open reading frames. Identified promoters are indicated as small black arrows, while the origin of replication is shown as a black box below the open reading frames. Regions of the genome transcribed early, middle, or late in the infection cycle are marked with black lines below the genome.

divergently located cluster of genes involved in lytic growth, whereas the virulent phages have similar organizations, only lacking the lysogenic gene cluster or parts of it (figure 36-2). Due to the similarity in genetic organization and the availability of many biological data from several different phages, the knowledge of P335 phages will be presented as functional modules rather than as isolated overviews of individual phages.

Transcriptional Patterns

Transcription of the temperate phage TP901-1 genome (figure 36-2) was analyzed by northern blot during the lytic cycle. It was found that sequential clusters of the genome were temporally transcribed and that the genome can be divided into early, middle, and late expressed regions (75). Eight early transcripts are present at maximum level 10 minutes after infection. At least seven of these originate from the two identified early promoters. The four middle transcripts, observed at maximum level 30 minutes after infection, are located in the region encoding *orf24* to *orf29*, which is also transcribed in the early phase of the lytic cycle. The remaining genes (*orf30* to *orf56*) are transcribed late in the lytic cycle (23, 75).

Site-Specific Recombination

Most phage integrases belong to the lambda family of integrases; only a few integrases are members of a new family showing homology to the catalytic site of resolvases and invertases (31, 64, 79, 109). TP901-1 encodes a resolvase-type integrase (group II in table 36-3), while the remainder of the investigated lactococcal phages contain an integrase belonging to the λ family (groups I and III in table 36-3). The TP901-1-encoded resolvase-type integrase catalyzes site-specific recombination in *L. lactis* (31), in *E. coli* (17), and in human cells (107).

Site-specific integration in *L. lactis* occurred into a single chromosomal site, *attB* (table 36-4), in both the host and the non-host strain MG1363 (31, 32). The frequency of lysogenization was determined to be 2–5% of the infected cells, using an erm-labeled phage (61). Furthermore, a TP901-1 mutant containing a deletion of part of the integrase gene and *orf2* showed the expected, dramatic reduction in lysogenization (61).

Integration vectors that are useful for study of promoter fusions in a single copy on the chromosome have been constructed based on nonreplicating plasmids containing *attP* from TP901-1 and a selectable marker. These integration vectors integrate very efficiently with a frequency similar to the frequency of plasmid transformation (21). The smallest regions of *attP* and *attB* that are sufficient for site-specific recombination were identified to 56 and 43 bp, respectively (17, 21) and the integrase binds with equal affinity to the two linear DNA fragments in vitro (17). A 5 bp sequence was identified as the core sequence (32) (table 36-4) and mutant analysis verified that the TC dinucleotide constitutes the overlap region (17).

Table 36-3 Similarity of Selected Functions Within the P335 Phages

Function	Group I	Group II	Group III	Group IV	Group V
Integrase	**r1t (ORF1)** bIL285 (ORF1: 98%) ΦLC3 (INT: 97%) BK5-T (ORF33: 97%) Tuc2009 (ORF1: 97%) bIL309 (ORF1: 40%) TPW22 (INT: 27%)	**TP901-1 (ORF1)**	**bIL286 (ORF1)** bIL309 (ORF1: 25%) ul36 (ORF359: 99%)		
Super infection exclusion	**r1t (ORF2)** Tuc2009 (SIE$_{2009}$: 100%) ΦLC3 (ORF173: 100%)	**TP901-1 (ORF2)** bIL285 (ORF2: 98%) bIL309 (ORF2: 66% C)	**BK5-T (ORF34)**		
Repressor	**r1t (RRO)** bIL309 (ORF3: 97%) ΦLC3 (ORF286: 79%) Tuc2009 (ORF4: 78%) BK5-T (ORF35: 79%)	**TP901-1 (ORF4)** bIL285 (ORF4: 98%) Φ31 (ORF180: 68% N) 4268 (ORF1: 67%N) ul36 (ORF188: 53%) bIL285 (ORF4: 32%)			
Cro-like	**r1t (TEC)** bIL309 (ORF7: 100%)	**TP901-1 (MOR)** bIL285 (ORF5: 98%) Φ31 (ORF74: 50%) ul36 (ORF74A: 50%) 4268 (ORF2: 50%)	**BK5-T (ORF37)**	**Tuc2009 (ORF5)** ΦLC3 (ORF76: 100%)	
Antirepressor	**r1t (ORF5)** BK5-T (ORF38: 96%) bIL309 (ORF8: 40% N)	**TP901-1 (ORF6)** bIL285 (ORF6: 100%) bIL286 (ORF6: 65% C) ΦLC3 (ORF236: 95% C) Tuc2009 (ORF6: 94% C) Φ31 (ORF238: 94% C) Φ31.1 (ORF238: 94% C) ul36 (ORF238: 94%C) 4268 (ORF3: 94%C)			
Excisionase	**r1t (ORF6)** ΦLC3 (ORF110: 99%) Tuc2009 (ORF7: 97%) bIL285 (ORF7: 97%) BK5-T (ORF39: 97%) ul36 (ORF111a: 97%) ul36.1 (ORF111b: 97%) 4268 (ORF4: 95%) bIL309 (ORF9: 75%)	**TP901-1 (ORF7)**			
Replication protein	**Tuc2009 (ORF16)** bIL285 (ORF16: 99%) ul36 (ORF255: 99%) r1t (ORF11: 30% N) bIL309 (ORF14: 54% C) 4268 (ORF11: 53%C) bIL286 (ORF16: 24% C)	**TP901-1 (ORF13)** BK5-T (ORF49: 27% C) ul36.1 (ORF235: 33% C) Φ31.1 (ORF269: 33% C)	**Φ31 (ORF492)**		
Holliday junction resolvase	**r1t (ORF14)** bIL285 (ORF18: 99%) ul36 (ORF129: 98%) Tuc2009 (ORF19: 97%) BK5-T (ORF51: 97%)	**TP901-1 (ORF15)** Φ31.1 (ORF139: 97%) ul36.1 (ORF109: 76%) bIL309 (ORF17: 32%)			
Large terminase subunit	**r1t (ORF27)** Φ31(ORF5: 97%) ΦLC3 (ORF180’: 97% N)	**TP901-1 (ORF31)** ul36 (ORF462: 97%) Tuc2009 (ORF31: 66% N) Tuc2009 (ORF32: 97% C)	**BK5-T (ORF2)** bIL286 (ORF41: 99%) 4268 (ORF31: 72%) bIL309 (ORF39: 34%) bIL285 (ORF41: 23%)		

(Continued)

Table 36-3 Similarity of Selected Functions Within the P335 Phages

Function	Group I	Group II	Group III	Group IV	Group V
Major head protein	**r1t (ORF31)**	**TP901-1 (ORF36)** Tuc2009 (ORF37-39: 98%)	**BK5-T (ORF7)** 4268 (ORF35: 97%) bIL286 (ORF45: 96%) bIL285 (ORF44: 23%)	**bIL309 (ORF43)** bIL285 (ORF44: 26%)	**ul36** (ORF287)
Lysin	**r1t (ORF49)**	**TP901-1 (ORF53)** ul36 (ORF429: 96%) ΦLC3 (LYSB: 95%) TPW22 (ORFA: 94%) Tuc2009 (ORF57: 93%)	**BK5-T (ORF27)** bIL285 (ORF62: 98%) bIL286 (ORF61: 97%) bIL309 (ORF56: 97%) Φ31(LYS: 98% N)	**4268 (ORF49)**	

For each group the identity of proteins encoding the same function is indicated as a percentage of the type protein encoded by the phage marked in bold. C, Similarity only in the C-terminal; N, similarity only in the N-terminal. Proteins showing high similarity are placed in the same group; however, some similarity may be found between groups, for example repressor groups I and II. In addition, some proteins may belong to the same protein family without being in the same group, for example integrase groups I and III that contain λ-like integrases.

Table 36-4 Attachment Sites of Temperate P335 Phages

Phage[a]	Size of core sequences Sequence 5′ to 3′ (if smaller than 15 bp)	Location of *attB* [b]	Reference to *attB*
ΦLC3, Tuc2009, r1t	9 bp TTCTTCATG	Noncoding (IL1403)[c]	(30)
BK5-T, bIL285		C-terminal of ORF (MG1363)[c]	(68)
bIL286	80 bp	Noncoding (IL1403)	(30)
bIL309	29 bp	tRNA^{Arg}, but intact gene preserved in lysogen (IL1403)	(30)
TPW22	14 bp TAAGGCGACGGTCG	Orf, disrupted in lysogen (ligase-like gene) (MG1363, 3107 and W22)	(94)
TP901-1[d]	5 bp TCAAT	Orf125, disrupted in lysogen (competence-like: *comGC* in *B. subtilis*) (MG1363 and 3107)	(17)

[a] All except TP901-1 contain λ-like integrases.

[b] Strains used for *attB* determination are indicated in parentheses.

[c] The core sequences for ΦLC3 and bIL285 were found to be located differently in the strains MG1363 (changing the last 5 amino acids in an ORF) and IL1403 (located after the homologous ORF, which was found to be shorter than in MG1363). Sequence information from MG1363 was obtained from M. Lunde.

[d] The overlap region was determined to TC (17).

Thus, mechanistically the integrase from TP901-1 so far behaves as other members of the resolvase family (105, 109).

Amongst the λ-type integrases almost identical proteins were found in phages ΦLC3, Tuc2009, r1t, bIL285, and BK5-T (table 36-3). Accordingly, identical *attP* and *attB* core regions of 9 bp have been determined for these phages (table 36-4). Site-specific integration catalyzed by the ΦLC3- and the Tuc2009-encoded integrases have furthermore been demonstrated in the non-host strain, MG1363 (68, 110). An isolated ΦLC3 *int* mutant could not form stable lysogens. Integrase activity, however, could be complemented in *trans* by a clear-plaque mutant of ΦLC3 (68). An integration vector based on the *attP* site and *int* from phage ΦLC3 has been constructed (70). The integration was highly efficient using a temperature-sensitive origin of replication. However, the number of integrants obtained using a nonreplicating plasmid suggests that the integration efficiency is in the order of 0.2%. A similarly low frequency of integration was obtained for the identical integration system of phage Tuc2009 (110). It is unsolved whether both systems contain insufficient promoters for integrase expression or whether the observed numbers instead reflect rather inefficient integration systems. The frequency of lysogenization by ΦLC3 has been estimated to be 0.01% of the cells by measuring the number of *attR* sequences by quantitative polymerase chain reaction (73). Furthermore, ΦLC3 prophage stability varied with temperature and growth phase of the host (74). However, the low frequency of lysogenization and the impact of the growth conditions of the host on prophage stability may also reflect properties of the genetic switch.

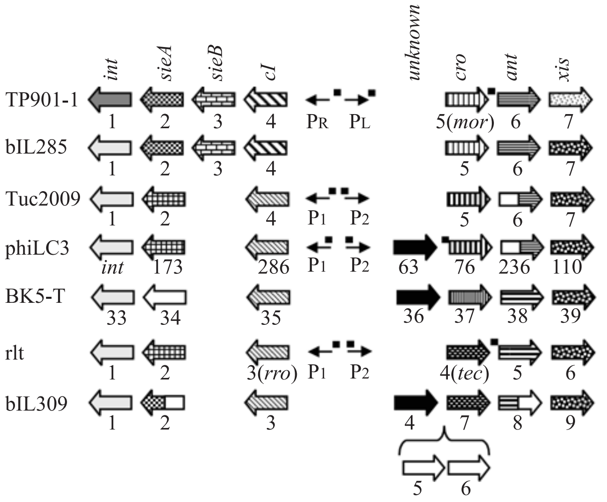


Figure 36-3 Genetic organization of the lysogeny module of temperate P335 phages. Identified and putative functions are indicated at the top of the figure and genes are aligned according to this. Integrase, *int*; superimmunity exclusion, *sieA*; part of Sie system and metalloproteinase motif protein, *sieB*; *cro*-like, *cro*; antirepressor, *ant*; excisionase, *xis.* Genes encoding similar proteins are shown as arrows with identical pattern. No similarity to other phage proteins, white; similar proteins with no identified function, black. Small black arrows indicate promoters identified by biological experiments, while locations of identified repressor operator sites are shown as small black boxes.

Table 36-4 summarizes the core regions and the locations of the *attB* sites for the lactococcal phages. The *attB* sites for several of the phages were localized in noncoding regions of the chromosome and hence no gene disruption occurs by lysogenization. Exceptions were TP901-1 and TPW22 lysogens that both contain a disrupted ORF and phage bIL309 that inserts into a $tRNA^{Arg}$ gene, resulting in preservation of an intact gene (table 36-4).

Excision of a TP901-1-based integration vector is dependent on a functional integrase. However, the frequency of excision is very low when the integrase is the only phage protein present. In contrast, 100% excision is found when ORF7, encoding the excisionase, is provided as along with the integrase (16). Excisionases have not been published for other lactococcal bacteriophages; however, a phage integrase and excisionase are expected to co-function during excision of the integrated prophage (as reviewed in chapters 7 and 8). Thus, a homologous integrase implies a homologous excisionase and comparison of proteins encoded by the early region of phages containing essentially the same integrase (r1t, Tuc2009, bIL285, BK5-T) revealed only one highly conserved protein present in all these phages (figure 36-3). We suggest that this small protein is the excisionase of these phages (r1t: ORF6; Tuc2009: ORF7; bIL285: ORF7; BK5-T: ORF39) (table 36-3). Data from S. Leach et al. (personal communication) confirm this with respect to ORF7 from Tuc2009. Phage bIL309, which contains an integrase 40% identical to the above-mentioned integrase, encodes a protein showing 75% identity to the proposed excisionase and it is therefore likely that this protein (ORF9) is the excisionase of phage bIL309 (figure 36-3, table 36-3).

The Genetic Switch

A characteristic feature of a genetic switch in temperate phages is a phage-encoded repressor able to repress early promoters controlling expression of genes necessary for phage development leading to the lytic cycle. Another feature is a factor counteracting repression, such as the Cro protein in phage λ which is able to repress expression of the CI repressor and hence is needed for the choice leading to lytic development (reviewed in chapter 8).

Two classes of repressors have so far been identified in lactococcal phages: the r1t and the TP901-1 classes (table 36-3). Sequence and folding analysis of repressors from both classes placed them in the HTH-3 family of DNA binding proteins due to a structure in the N-terminal end of the repressor protein. The C-terminal ends are, in analogy with CI from λ, presumed to harbor the regions responsible for repressor subunit–subunit interactions. The difference between the two classes of lactococcal repressors is their size (278 and 180 amino acids for r1t and TP901-1, respectively), as well as the lack of sequences for RecA-mediated autodigestion in the TP901-1 class of repressors (figure 36-4). It should be mentioned that all characterized temperate lactococcal phages nevertheless are inducible by mitomycin C, including those having the TP901-1 type of repressor.

Comparisons of the amino acid sequences of the r1t-like repressors (r1t, bIL309, ΦLC3, Tuc2009, and also BK5-T; table 36-3) revealed that approximately two thirds of the total amino acid sequence was identical, while less sequence similarity was found in the N-terminal domain. The TP901-1-like repressors show approximately 33–36% identity to the phage r1t-like repressors but only in a short region (figure 36-4). This region might be involved in protein–protein interactions with host proteins, for example RecA. A characteristic feature is the distance between the start sites of the lytic and lysogenic promoters (approximately 100 bp) and the presence of two operator sites in the promoter region and often also one operator site overlapping the first gene of the lytic operon (figure 36-3).

The repressor Rro from phage r1t was the first lactococcal phage repressor to be analyzed. Three palindromic operator sites were identified by gel retardation studies. Primer extensions identified two promoters: P1 and P2 (figure 36-3) (87). The expression from the lytic P2 promoter was studied using a translational fusion to *lacZ*. In the fusion plasmid containing both *rro* and *tec*, P2 expression was shown to be inducible by mitomycin C in the non-host

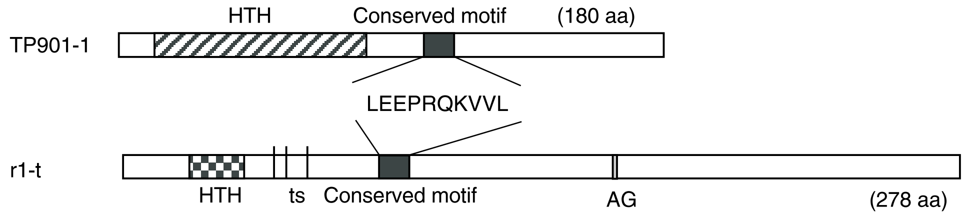


Figure 36-4 Comparison of repressors of P335 temperate phages TP901-1 and r1t. Position of helix–turn–helix (HTH) motifs representing putative DNA binding domain are indicated with different shading in TP901-1 and r1t, while the conserved motif found in all lactococcal phage repressors is indicated as a black box. For r1t the position of the mutations leading to a temperature-sensitive repressor (ts) and the putative signature sequence for RecA-catalyzed autodigestion (AG) are also indicated.

strain *L. lactis* ssp. *cremoris* MG1363, thus mimicking the behavior of the prophage. Introduction of a 5′-located frameshift mutation in *rro* was expected to result in constitutive expression of the lytic promoter. However, gradually more and more derepressed P2 expression during growth of the bacterium was observed and this phenomenon was not explained. Temperature-sensitive mutants in the Rro repressor were isolated by random mutagenesis of three amino acids — F50, V54, and P61 — which were selected after folding of the N-terminus of Rro after the phage λ CI structure and using the information from the temperature-sensitive A66T mutation in this protein (presumed to correspond to V54 in Rro) (86). These results strongly suggest that the DNA binding domain of Rro from phage r1t, as in λ, is composed of a helix–turn–helix motif in the N-terminus of the protein. The repressor protein encoded by bIL309 is nearly identical to Rro, suggesting that r1t and bIL309 are homoimmune phages.

An isolated clear-plaque mutant of phage ΦLC3 contains an amber mutation in *orf286*, suggesting that the gene encodes the immunity repressor. However, this has not been further investigated (10, 69). ORF286 of ΦLC3 belongs to the r1t-type of repressors, which are almost identical to ORF4 of Tuc2009 except for four amino acids. The two phages ΦLC3 and Tuc2009 are therefore proposed to be homoimmune (10).

Activity of two divergent promoters, P1 and P2 in ΦLC3, was confirmed by the use of transcriptional fusions, and transcriptional start sites were identified by primer extension analysis (10, 11). Two operator sites overlapping both promoters were identified by footprint analysis, while binding to a third operator site located 500 bp downstream in the lytic operon was shown by gel retardation (10) (figure 36-3). The binding region contains only imperfect inverted repeats with different spacing. However, a direct-repeat sequence, CGTGGTT, was also identified (10). The proposed asymmetric operator sites were found for the almost identical CI from phage Tuc2009 and alignment of the promoter regions of the two phages shows considerable identity (66).

In vitro experiments predict that the repressor ORF286 of ΦLC3 can repress both promoters, P1 and P2. However, only a 2-fold repression of the lysogenic promoter P1 was observed in vivo (10). Interestingly, the activity of the P1 promoter increased 10-fold when both repressor and the Cro-like protein, ORF76, were present and it was suggested that ORF286 is able to stimulate its own synthesis at low concentrations (11). Furthermore, the results suggest that ORF76 is able to repress the activity of both the lytic and lysogenic promoters and it was demonstrated that ORF76 binds specifically to the genetic switch region, albeit with a lower affinity than the repressor (11). The operator sites for ORF76 binding could not be determined based on gel retardation experiments and the authors suggest that the repressor (ORF286) and the Cro-like protein (ORF76) compete for DNA binding (11).

By the use of a plasmid-borne translational fusion it was shown that the lytic promoter (P2) of Tuc2009 is induced by addition of mitomycin C in the presence of the CI and Cro proteins. Introduction of a nonsense mutation in *cI* did not result in constitutive P2 expression. Instead, a gradual activation was seen, similar to results obtained with phage r1t (87). His-tagged Cro protein binds to the intergenic region and, in contrast to the results obtained with the similar ORF76 protein encoded by ΦLC3, it is proposed that Cro of Tuc2009 functions as an activator for the lytic promoter (D. van Sinderen, personal communication) Further experimentation is required to clarify the role of the Cro-like proteins in regulation of the genetic switch of phages ΦLC3 and Tuc2009.

The CI proteins from phages r1t and Tuc2009 differ only in their N-terminal region. A helix-swap experiment was performed (D. van Sinderen, personal communication) resulting in a Tuc2009 repressor that binds to r1t operator sites but not to Tuc2009 operators. This finally identifies the specificity of the DNA binding to the helix region KTTISNYEV of Tuc2009. It was also shown that His-tagged repressor proteins from both r1t and Tuc2009 give immunity to the corresponding phage. However, the phages are heteroimmune as expected (D. van Sinderen, personal communication).

ORF37 encoded by the second gene in the lytic operon of phage BK5-T was proposed to be involved in the genetic switch, since it seems to bind to the promoter region and also contains a putative helix–turn–helix DNA-binding motif. However, ORF37 does not show similarity to proteins encoded by P335 phages, but it shows similarity to topological Cro proteins in *Streptococcus thermophilus* phages (78, 88).

In phage TP901-1 two divergently located promoters, P_R (lysogenic) and P_L (lytic), were identified by primer extension analysis and promoter fusions (75, 76). The phage repressor (CI), encoded by *orf4*, was shown to be necessary for repression of both promoters and to confer immunity of the host strain to TP901-1 infections (76). Two clear-plaque mutants of TP901-1 have been analyzed, both containing mutations in the ribosome binding site for CI that reduce CI expression from the phage (K. Hammer, unpublished results). The CI repressor was shown to bind in vitro to a 317 bp DNA fragment covering P_L and P_R, and the binding was found to be cooperative (56). DNase I footprinting identified two palindromic binding sites, O_L and O_R (figure 36-3). Three operator mutants selected as having increased expression from the lytic promoter, P_L, were all found to be located in O_L, confirming the major role of O_L in repression of the lytic promoter. A third O site, O_D, located downstream of *mor*, was identified by gel retardation (56). The purified repressor was found to exist in solution both as a dimer and as multimeric forms (hexamers or higher), suggesting that multimeric forms of the repressor could be involved in cooperative binding to all three operator sites. The location of the operator sites results in more repression of the lytic promoter, P_L, than of the lysogenic promoter, P_R, during cooperative binding of the repressor.

In phage TP901-1 the first gene in the lytic operon (*mor* for modulator of repression) was found to influence the repression of the lytic promoter, P_L, and the results suggested that the relative amounts of CI and MOR after phage infection determine the decision between a lytic or lysogenic life cycle (76). For instance, when *L. lactis* ssp. *cremoris* is transformed with a plasmid containing the two divergent promoters from TP901-1 (P_L and P_R), *cI* and *mor* clonal variation is observed, since P_L is either active or repressed in each transformant. The clonal variation required the presence of both *cI* and *mor*. The repression of the promoters was still dependent on the repressor since MOR alone had no effect on the activity of the P_L promoter. However, a surplus of MOR was able to relieve CI repression; conversely when P_L was derepressed, a surplus of the CI repressor resulted in repression of the promoter even though MOR was present. Furthermore, mitomycin C induction of the lytic P_L promoter required the presence of both CI and MOR. CI-MOR protein interactions thus have been suggested to explain the data for this TP901-1 switch (76). Phage bIL285 encodes nearly identical repressor, operator sites, and MOR proteins suggesting that bIL285 and TP901-1 are homoimmune and furthermore that the genetic switch is identical in the two phages. The temperate phages ΦPVL and Φ1205 from *Staphylococcus aureus* and *Streptococcus thermophilus*, respectively, also show high similarity to TP901-1 indicating the same kind of genetic switch also in these phages (56).

Surprisingly, the virulent phages Φ31 and ul36 were found to encode CI homologs (65, 77) which show some similarity to the N-terminal of the TP901-1 repressor (table 36-3). By the use of plasmid-borne transcriptional fusions in the non-host strain, MG1363, it was shown that the presence of the *cI* gene of Φ31 does not repress transcriptional initiation of the lytic promoter, P1, whereas expression from the lysogenic promoter, P2, was repressed 2-fold. In contrast, the presence of the first gene (*cro*) in the lytic gene cluster of Φ31 efficiently represses both promoters (77), which may represent the ability of CRO to shut down early gene expression during infection. It seems likely, therefore, that CI has lost the ability to repress the lytic phage promoter. To evolve into a potent virulent phage, Φ31 may furthermore have accumulated mutations in the *cI* gene giving a negative dominant phenotype in the presence of an active repressor from the host chromosome. The finding that small C-terminal deletions in CI re-establish the ability of the mutated CI to re-press the Φ31 lytic promoter supports this hypothesis (38). The truncated CI protein was further shown to repress 10 different virulent phages, demonstrating a whole family of virulent phages which have probably evolved from a temperate phage having a TP901-1-like repressor. Phage mutants that were resistant to the repression from the truncated CI repressor identified palindromic operator sites overlapping the lytic promoter and at the end of *cro* (38). By sequence comparison a third operator site was located between the −35 regions of the two divergent promoters. The location of the three operator sites matched exactly the location identified for the operator sites in TP901-1, and the operator sequences also showed considerable similarity. The CRO protein of Φ31 showed only 50% identity to the MOR protein in TP901-1. The protein, however, retains a different function since it is able to repress both of the early phage promoters by itself.

The Cro-like protein is not encoded as the first gene in the lytic operon in all temperate P335 phages (figure 36-3). In phage ΦLC3, ORF63 is identical to the corresponding *orf36* in phage BK5-T and ORF4 of phage bIL309 (figure 36-3), while some similarities (50%) to ORF8 of phage r1t and ORF9 of phage TP901-1 were found. It has been reported that these proteins of phages ΦLC3 and BK5-T do not bind to promoter-containing DNA in vitro, suggesting that they may not be involved in the genetic switch (10, 78). Future research may clarify the role (if any) of the genes preceding the Cro-like gene in the regulation of the genetic switch in temperate P335 phages.

Superinfection Exclusion

Superinfection exclusion (Sie) functions are expressed in the prophage and confer immunity of the host cell to heterologous superinfecting phages; they are not involved in maintenance of the lysogenic state. Recently the superinfection exclusion gene (*sie*$_{2009}$) has been identified in phage Tuc2009 (81). Sie$_{2009}$ is associated with the membrane and is predicted to contain a transmembrane helix. Expression of *sie*$_{2009}$ in *L. lactis* provides little or no protection against phages of P335 or c2 species, whereas complete resistance against some members of the 936 phage species was obtained (81). The authors suggest that phage proliferation was prevented by injection blocking since cells expressing Sie2009 allow phage adsorption but not transduction. The P335 phages r1t and ΦLC3 encode proteins highly similar to Sie$_{2009}$ (figure 36-3, table 36-3), and in ΦLC3 lysogens high levels of mRNA were detected from this gene (*orf173*) (11). ORF2 of phage bIL309 shows similarity in the C-terminal end to Sie$_{409}$ encoded by a prophage of *L. lactis*, IL409, that was found to mediate a phage-resistance phenotype similar to Sie$_{2009}$ (81). A prophage in *L. lactis* F7/2 encodes another superinfection exclusion system consisting of two genes (*sie*$_{F7/2A}$ and *sie*$_{F7/2B}$), which both must be expressed to obtain a phage resistance phenotype for three 936-type phages tested. Sie$_{\mathrm{F7/2A}}$ and Sie$_{\mathrm{F7/2B}}$ are identical to ORF2 and ORF3 of phage TP901-1 and show high similarity to ORF2 and ORF3 of bIL285 (81). In accordance with this, *orf2* and *orf3* of phage TP901-1 are expressed in a strain lysogenic for TP901-1 (75). ORF34 of phage BK5-T (figure 36-3) is expressed in the prophage and it was suggested that these proteins might have a role in superinfection exclusion, although no evidence for this function was obtained (14).

Antirepressor and Genetic Organization of the Lysogeny Module

In *E. coli* phage P1 and *Salmonella* phage P22 (reviewed in chapters 24 and 29, respectively), the antirepressor Ant destroys repressor activity by protein–protein interactions. So far no clear evidence for such protein functions in phages infecting lactic acid bacteria has been presented. The MOR protein from TP901-1 is suggested to function as an antiprepressor (76) based on the in vivo results, and the Ant homolog (ORF6) from Tuc2009 has been shown in vitro to counteract DNA binding of the Cro protein, which has been proposed to function as a phage activator (S. Leach, personal communication). In the streptococcal phage, Sfi21, the Cro protein has also been shown in vitro to inhibit DNA binding of the repressor (20). In most P335 phages Ant proteins have been suggested based on partial similarity to the Ant proteins of phages P1 or P22 or other phage proteins located downstream of the *cro*-like gene. The putative antirepressors of P335 phages contain a two-domain structure (72), and, based on amino acid similarity, two classes of antirepressors are observed (table 36-3). No significant similarity was found between the two classes.

The genetic organization of the lysogeny module of *Sipoviridae* phages infecting low G-C content Gram-positive bacteria is highly conserved (72) with the following gene order: integrase, superinfection exclusion, in some cases a metalloproteinase motif gene, repressor, and a *cro*-like gene followed by a putative antirepressor (figure 36-3). However, some phages contain insertions of up to three genes upstream of the *cro*-like gene (figure 36-3). The recent results from the P335 phages show that the metalloproteinase is part of a two-gene superinfection exclusion system in some phages (81). Furthermore, our comparative analysis identified the excisionase gene downstream of the putative antirepressor in all the sequenced temperate P335 phages including BK5-T. Downstream of the excisionase gene the temperate P335 phage contains a variable number of genes, which currently are of unknown function. However, in phage TP901-1 we have observed that transcription from the early lytic promoter, P_L, is repressed during the later stages of infection and that some of these genes with otherwise unknown functions may potentially be involved in this process (K. Hammer, unpublished data).

DNA Replication

Even though bacteriophages require many bacteria-encoded proteins for replication, they often, in addition to the phage origin of replication, contain a genomic region that specifies a protein used for sequence-specific initiation of replication and one or more proteins involved in replication initiation. A phage encoded resistance (Per) phenotype has been described among the P335 phages as resistance against phage infection by hosts containing the phage origin of replication in multiple copies. Titration of proteins involved in DNA replication has been suggested to account for this phenomenon (49).

In bacteriophage Tuc2009 a 160 bp internal region in *orf16*, containing four direct repeats, inhibited DNA replication (and hence lytic growth of Tuc2009) when present on a multi-copy plasmid. A library of mutations in this fragment (*ori*$_{2009}$) was analyzed and most of the mutations that abolished the inhibitory effect on proliferation of phage Tuc2009 were localized in the direct repeats (82). Using gel retardation analysis, ORF16 was shown to be able to bind to *ori*$_{2009}$ and ORF16 therefore was proposed to be the replisome organizer of Tuc2009, now called Rep$_{2009}$. Interestingly, Rep$_{2009}$ could still retard the DNA fragments harboring the isolated Tuc2009 proliferation-inhibiting mutations and the authors propose that another protein, in addition to Rep$_{2009}$, may bind to the *ori*$_{2009}$ region (82). Recently, it was shown that strains producing antisense mRNA of *rep*$_{2009}$ and *orf17* acquired a very pronounced resistance against Tuc2009 and a reduction in internal phage DNA replication, indicating that the protein product of *orf17* very likely is involved in Tuc2009 replication (80).

A Per phenotype of ori_{2009} was observed for phages Q30, Q33, and ul36 (80) and the presence of the putative replication protein and origin of phage ul36 were similarly shown to inhibit proliferation of phage ul36 and Q33 (13). This verifies that these origins are functional homologs. In addition, sequence analysis of phage ul36 (figure 36-5) and hybridization studies of phages Q30, Q33, and ul36 showed regions that are highly similar to ones found in Tuc2009 (*orf15* to *orf19*). However, *orf18* of Tuc2009 encoding a putative type II methyltransferase was not present in these phages (13, 80). Two strains producing antisense mRNA of the Tuc2009 genes rep_{2009} and *orf17* showed resistance to the phages Q30, Q33, and ul36, which strongly suggests that both rep_{2009} and *orf17* are involved in DNA replication in these phages (80).

The protein product of *orf11* (Pro11) of phage r1t was shown to bind specifically to sequences located within *orf11* and footprinting revealed that the protected DNA region contained four 6 bp direct repeats, suggesting that these repeats constitute the *ori* of phage r1t (119). No in vivo experiments have been provided. However, the Pro11 protein shows similarity to the replication proteins G38P of *B. subtilis* phage SPP1, indicating that Pro11 may be involved in r1t DNA replication (phage SPP1 is reviewed in chapter 23).

Within *orf13* of phage TP901-1, several repeats were located and the presence of these repeats on a plasmid *in trans* was shown to confer Per resistance of the host strain against TP901-1 infection (120). TP901-1 prophages that contain mutations in the replication region were constructed. One prophage mutant contained mutations in the repeats without any change of the amino acid sequence of ORF13, while an amber stop codon in *orf13* was introduced in another. When the mutated prophages were induced, internal phage DNA replication was absent and the number of plaque-forming phages was reduced 10^6-fold compared with the wild-type. This supports the hypothesis that the repeats are the origin of replication of TP901-1. Furthermore, the effect of the amber mutation in *orf13* on phage proliferation and internal phage DNA replication could be relieved by the presence of an amber suppressor, showing that ORF13 is involved in TP901-1 DNA replication (120).

A Per phenotype identified putative highly identical origins of replication in the virulent phages Φ31.1 and ul36.1 (13, 37). ORF49 of BK5-T showed significant similarity to the putative replication proteins of phages Φ31.1 and ul36 (ORF269 and ORF235, respectively), and a repeat-rich region within *orf49* was shown to confer phage resistance (78). In the virulent phage Φ31 an AT-rich region present on a high-copy-number plasmid reduced the efficiency of plating of phage Φ31 by 10^{-6}, indicating that this region may carry the origin of replication (77). The Φ31 origin of replication was located downstream of a putative primase gene and its position in a noncoding region is unique among the P335 phages.

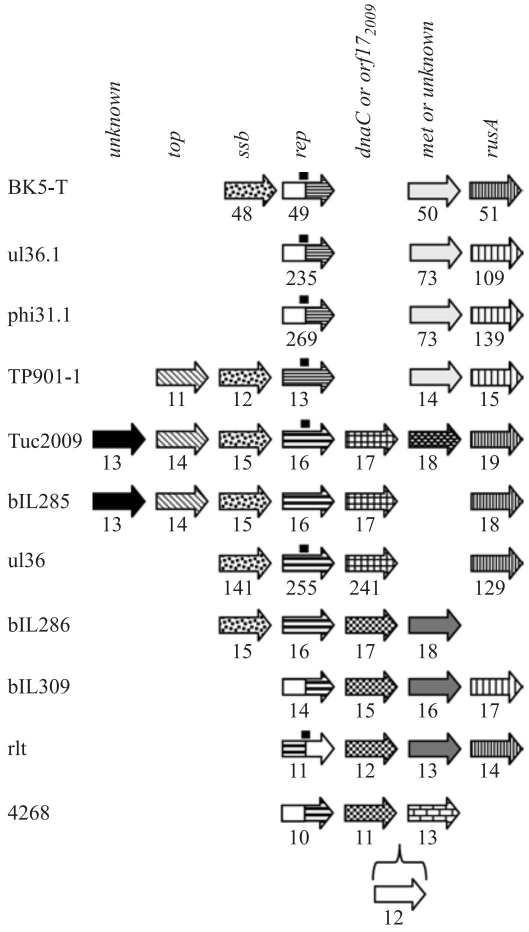


Figure 36-5 Genetic organization of the replication region of P335 phages. Identified and putative functions are indicated at the top of the figure and genes are aligned according to this. Putative topoisomerases, *top*; single-stranded binding proteins, *ssb*; replication initiation proteins, *rep*; homologs to ORF17 of Tuc2009, $\mathit{orf17}_{2009}$; DnaC homologous proteins, *dnaC*; methylase, *met*; putative Holliday junction resolvases, *rusA*. Genes encoding similar proteins are shown as arrows with identical pattern. No similarity to other phage proteins, white, similar proteins with no identified function, black or gray. Small, black boxes indicate the location of origin of replication identified by experiments.

The replication proteins of P335 phages can be divided into several similarity groups (table 36-3). One contains the TP901-1-like replication proteins among which the replication proteins of phages BK5-T, ul36.1, and Φ31.1 are highly similar. Computational analysis suggested that the replication protein of phage TP901-1 might be divided into two domains linked by a short hinge region that encodes the origin of replication (36). The replication proteins of phages Tuc2009, ul36, and bIL285 are highly similar, while the replication proteins of phages bIL309, 4268, r1t, and bIL286 show some similarity to the Tuc2009-like replication proteins (figure 36-5, table 36-3). However, these homologies are restricted to either the N-terminus (r1t) or the C-terminus (bIL286, bIL309, and 4268) of the protein, indicating a two-domain structure as suggested by (36). The putative replication protein of phage Φ31 does not show any similarity to either group of lactococcal phage-encoded replication proteins. However, some similarity to *S. thermophilus* phages was found (77).

Based on amino acid similarity several proteins have been suggested to be involved in phage DNA replication (figure 36-5). These are putative topoisomerases, single-stranded binding proteins, DnaC homologous proteins, Holliday junction resolvases, and dUTPases. The function of the putative topoisomerase and single-stranded binding protein does not seem to require corresponding similar replication proteins, suggesting that no specific protein interaction takes place during phage DNA replication (compare TP901-1 and Tuc2009 in figure 36-5). The r1t RusA protein encoded by *orf14* was shown to resolve Holliday junctions in vitro, and it furthermore promotes DNA repair in *E. coli* strains lacking the RuvABC resolvase (103). The putative Holliday junction resolvase encoded by many P335 phages may thus possibly resolve branched DNA structures formed during DNA replication or alternatively during DNA packaging. Two different types of Holliday junction resolvases are found. However, the presence of a specific type seems to be independent of the sequence of the replication protein. Interestingly, all P335 phages encode nearly identical dUTPases, which are significantly different from the chromosomally encoded protein in *L. lactis* IL1403.

Regulation of Middle and Late Gene Expression

In the virulent bacteriophage Φ31 a promoter has been identified that is induced 20 minutes after infection, that is the middle phase of infection as defined by the authors (114). Transcription from this promoter is induced by the presence of a phage Φ31-encoded activator located upstream of the promoter on the Φ31 genome (89, 114). The activator of Φ31 was suggested to bind to sequences overlapping the −35 region of the promoter and deletion analysis showed that removal of the region from −54 to −45 eliminated promoter activity (114). The promoter and corresponding activator are conserved between phage Φ31 and the two temperate bacteriophages, r1t and ΦLC3 (113).

A TP901-1 promoter, active 30–40 minutes after infection, has been identified. It was designated the late promoter of phage TP901-1 since it becomes fully active in the late phase of the lytic cycle and controls transcription of the late-expressed region of the TP901-1 genome (22). This promoter is located upstream of the terminase subunits, a location that corresponds to the middle promoter of Φ31. Thus, the two promoters may be functionally equivalent. The TP901-1 promoter is tightly regulated and requires ORF29 for activity. The transcriptional start site of the promoter was identified by primer extension and is located in the intergenic region between *orf29* and *orf30.* The region located −85 to −61 bp upstream of the transcriptional start site was shown to be necessary for promoter activity and a region from −79 to −32 was found to be protected by ORF29 using footprinting analysis; this region contains four direct repeats (92). Similarity searches revealed that an ORF29 homolog and a sequence similar to the late promoter region are present in several genomes of temperate lactococcal bacteriophages (high similarity: Tuc2009, bIL286, ul36, bIL309; low similarity: 4268 and bIL285) (22).

DNA Packaging

The existence of cohesive DNA ends and hence *cos* sites has been demonstrated for the phages ΦLC3 and BK5-T (69, 78, 111). Identical *cos* sites (13 bp single-stranded 3′ overhangs with the sequence: 5′-GTGACGGCGTGAA-3′) were determined for phages ΦLC3 and r1t (69, 111) while the BK5-T *cos* site was identified as a 12 bp 3′ single-stranded overhang with the sequence 5′-CACACACATAGG-3′. The latter sequence was furthermore found upstream of *orf40* in the bIL286 genome, indicating that packaging of the bIL286 genome is initiated at this *cos* site. By sequence similarity the large subunit of the terminases could be identified in the *cos*-site P335 phages, and the terminases of phages BK5-T, bIL286, bIL285, and bIL309 show similarity (table 36-3). The N-terminus of a partly sequenced ORF encoded by ΦLC3 is highly similar to the large subunit of the terminase of phage r1t. This protein shares only limited similarity with other terminase subunits of lactococcal phages. In contrast, the protein shows similarity to terminases of *Siphoviridae* phages infecting high G-C content Gram-positive bacteria.

DNA packaging of temperate phage TP901-1 and Tuc2009 was suggested to be initiated at a *pac* site located upstream of *orf30* of TP901-1 and in a region upstream of *orf39* on the Tuc2009 genome (4, 23, 57). Based on protein similarity, protein domains, and location, *orf30* and *orf31* of TP901-1 were suggested to encode the small and large subunits of the terminase (23). ORF31 and ORF32 of Tuc2009 show similarity to the N- and C-terminal of the large terminase subunit of TP901-1, respectively, indicating that ORF31 and ORF32 of Tuc2009 may encode separated domains of the large terminase subunit.

Morphogenesis

Structural phage proteins have often been identified by SDS-PAGE and N-terminal sequencing, followed by analysis of the phage DNA sequence. SDS-PAGE investigations of r1t phage particles showed the presence of two very large protein bands of 160 and 190 kDa each (111). They both had an N-terminus corresponding to ORF31, strongly indicating that these proteins are covalently linked head proteins as observed for mycobacteriophage L5 and coliphage HK97 (46, 96). Another very abundant protein, ORF37, is probably the major tail protein (111). Less abundant protein bands corresponding to ORF27 and ORF45 were also observed. Similarity searches reveal that ORF27 may encode the large subunit of the terminase, whereas ORF45 shows weak similarity to the baseplate protein of phage TP901-1. A self-splicing group I intron seems to be contained within ORF41 (85).

Phages TP901-1, Tuc2009, and ul36 show extensive similarity in the genomic region encoding the structural proteins. The major differences are found in the baseplate proteins, the proteins specifying the neck passage structure (which is not found in ul36), and the major head proteins. In TP901-1 the major head protein (ORF36) was identified by SDS-PAGE and immunogold electron microscopy (58, 59). It is similar to the major head protein of Tuc2009 (major protein 2) encoded by *orf37* and *orf39*, which is suggested to be linked after excision of *orf38* that specifies a group I intron. Phage ul36 encodes a major head protein that is different from all other P335 phages (65). The major tail protein and a baseplate component were identified in TP901-1 (59, 93). They are similar to proteins identified as major structural proteins 2 and 1 of Tuc2009, encoded by *orf44* and *orf53*, respectively (4, 30). A tail component that radiates from the connection between head and tail called a neck passage structure was identified in TP901-1 (59). The neck passage structure (ORF51) shows high similarity to ORF47 from phage r1t and to ORFs encoded by phages belonging to the lactococcal 936 phage group. It has been shown that ORF45 functions as a tape measure protein involved in tail length determination, since introduction of an in-frame deletion or duplication in *orf45* of TP901-1 shortens or lengthens the TP901-1 tail, respectively. Furthermore, it was shown that ORF45 is important for assembly of the TP901-1 tail and that the baseplate and tail structures of TP901-1 assemble through a branched baseplate and tail assembly pathway (93).

In phage BK5-T, two major structural proteins were identified and N-terminal sequence analysis determined that they were encoded by *orf7* and *orf12*, respectively (78). The major head protein, encoded by *orf7*, was shown to be subject to proteolytic cleavage, since the protein product identified as the major head gene starts at amino acid number 110. The BK5-T protease responsible for the processing is most likely the ClpP like protease, encoded by *orf6*. The other major structural protein, encoded by *orf12*, is probably the major tail protein (78). The major head protein of BK5-T is nearly identical to the putative major head protein of phage bIL286 and 4268, whereas some similarity was found to the major head protein of bIL285 (table 36-3). Phages bIL285, bIL286, bIL309, and 4268 furthermore encode a protease, suggesting that the major head proteins of these phages also are subject to proteolytic cleavage.

Cell Lysis

Although some bacteriophages have evolved alternative systems for cell lysis, a generalized model involves the action of two principal proteins, holin and lysin (see chapter 10 for a general review of phage lysis). According to this model the holin molecules mediate the transport of the lysin protein across the cytoplasmic membrane by creating pores in the membrane. Subsequently, lysin degrades the bacterial cell wall from the outside. The P335 phages encode two types of murein-degradation enzymes (lysins) that attack either the glycosidic linkages between the amino sugars of the peptidoglycan (called muramidases) or the N-acetylmuramoyl-L-alanine amide linkage between the glycan strand and the crosslinking peptide (called amidases).

The lysin of phages Tuc2009 and ΦLC3 has been shown to have lytic activity against *L. lactis* and was suggested to encode muraminidase activity (4, 8). Lysin and holin proteins of phages Tuc2009, ΦLC3, and TP901-1 were highly similar (table 36-3). ORF48 and ORF49 of phage r1t has been suggested to encode a holin and a lysin with amidase activity, respectively (111). Also phage BK5-T was suggested to encode a lysin (ORF27) with amidase activity, while ORF25 and ORF26 of this phage have been proposed to encode a two-component holin system similar to that suggested for phages infecting *S. thermophilus* (78, 104). Phages bIL285, bIL286, and bIL309 each encode a holin and lysin similar to those encoded by BK5-T (table 36-3).

Concluding Remarks

Revised Taxonomy

DNA hybridization studies indicated that BK5-T and a few other uncharacterized temperate phages were not related to the type phage P335 (15), and BK5-T was therefore not originally placed in the P335 species (55). However, in the early region of the BK5-T genome identity was found to several P335 phages. In addition, BK5-T and the phages bIL285, bIL286, and bIL309 show protein similarity in the structural gene clusters exemplified by the large terminase subunit and major head protein (table 36-3), suggesting that these five phages belong to the same taxonomic group, namely the P335 phage species. The P335 phage species could, however, be divided into three different groups based on the protein similarity in the structural proteins as shown in table 36-5. This new classification of the lactococcal P335 phages is in agreement with the taxonomy suggested in (30, 35, 65). In contrast to the

Table 36-5 Revised Taxonomy of Lactococcal P335 Phages

Phage type	Prototype phage	Other lactococcal members	Related phages attacking different hosts (in parentheses)[a]
B1; P335 species: *cos*	r1t	ΦLC3 Φ31 (L)	TM4 (*Mycobacterium*) SF370.3 (d) (*S. pyogenes*)
B1; P335 species: *pac*	TP901-1	Tuc2009 ul36 (L)	Sfi11 and O1205 (*S. thermophilus*) Φgle and LL-H (*Lactobacillus*) A118 (*Listeria monocytogenes*) SF370.2 (d) (*S. pyogenes*)
B1; P335 species: *cos*	BK5-T	bIL285 bIL286 bIL309 4268 (L)	Sfi21, Sfi19, 7201, DT1 (*S. thermophilus*) Φadh (*Lactobacillus*)

[a] *S. Streptococcus*; L, lytic phage; d, defective phage.

prolate-headed c2 phages or the small isometric-headed 936 phages that contain highly similar genomes (more than 90% nucleotide identity), the P335 phages show lower and variable levels of DNA similarity (5–50%). BK5-T and bIL286 show extensive DNA identity in the structural gene cluster, and 45% overall identity, while about three quarters of the genomes of TP901-1 and Tuc2009 could be aligned (36). Furthermore, the virulent phage ul36 shows extensive identity to both TP901-1 and Tuc2009 (65).

Phages infecting different host genera also encode structural proteins that are similar to the lactococcal P335 phages, suggesting that these phages are related and could be placed in the same taxonomic group. However, in only a few cases could identity at the nucleotide level be detected between phages infecting different host genera. This was found between phages BK5-T/bIL286/bIL309 and *S. thermophilus* phage Sfi21, respectively (30, 35), and also for phage r1t and *Streptococcus pyogenes* phage SF370.3 (36). A revised taxonomy of the *Siphoviridae* has been proposed using comparative genomics (19). In this proposal the fully sequenced lactococcal phages of the phage species c2, 936, and P335 were placed into five genera. As described above, members of the P335 group, including BK5-T, do constitute three new genera. Furthermore, with the exception of the c2-like genus, the four remaining lactococcal genera were proposed to be members of a λ supergroup within *Siphoviridae* since they preserved the λ-like gene order in the structural genes (see chapter 27 for review of lambdoid phage comparative biology).

Evolution

The existence of two types of lactococcal phage populations has been proposed (30). One type consists of virulent phages with highly similar genomes (the c2 and 936 phage species), whereas the other corresponds to temperate phages showing very different levels of DNA similarity (the P335 phage species). The virulent P335 phages may all originate from temperate phages by deletion events. Hence, virulent phage ul36 contains remnants of a lysogeny module including a putative integrase gene (65), while phages Φ31 and 4268 contain a putative repressor but neither a phage attachment site nor an integrase gene has been identified (77; AF489521). However, the virulent phages of the P335 species may in principle return to the temperate status by acquiring the deleted part of the lysogenic module from related temperate phages. This cannot be achieved by virulent phages of the c2 and 936 species, since they do not have any temperate counterpart (30).

Many examples of horizontal gene transfer amongst the temperate phages may be suggested on the basis of genome comparisons (e.g., see chapters chapters 4 and 27), but biological experiments have also shown this to be the case. Phage Φ50, which is completely resistant to a plasmid-encoded R/M system, was found to encode a functional domain of the methylase gene identical to the plasmid-encoded methylase. This strongly indicates a recent genetic exchange between the phage and the plasmid (50). Mutant phages of Q30, Q33, and ul36 were isolated when the corresponding wild-type phage was plated on strains containing a plasmid with multiple copies of *ori2009*, normally conferring a phage-resistance phenotype. Two of three analyzed phages showed genomic reorganizations in the DNA replication region, while the third may harbor point mutations (80). Similar results were obtained for phage Φ31 (37). In both cases it was speculated that the recombinant phages have acquired new DNA replication regions by homologous recombination with prophage or prophage remnants in the host chromosome. Also the presence of an abortive resistance mechanism, *abiK*, that was shown to interfere with phage DNA replication resulted in the occurrence of recombinant phages after challenge with phage ul36 (13). These authors moreover showed that the sequence obtained by the recombinant phages originated from the host chromosome and defined the point of exchange between the chromosome and the incoming phage (13). More generally, it has been suggested that short regions of micro-similarity located in noncoding regions of the P335 phage genomes may be points of exchange between phages (23, 25).

The presence of prophages in the chromosome of the host bacterium increases the probability for two phage genomes being present in the same cell at the same time, in contrast to a simultaneous infection by two different phages. Furthermore, the life cycle of virulent phages depends on frequent and productive cycles of infection, thus greatly limiting the flexibility of incorporation of extra DNA except for genes conferring an immediate selective advantage. Thus, two modes of evolution of phages have been proposed: the temperate/lytic mode and the virulent mode (30, 48). Extensive horizontal gene transfer, the frequency depending on the relatedness of the bacterial hosts, characterizes the temperate/lytic mode. The virulent mode shows very little DNA exchange with other phage species, with the exception of host-related functions. An example of this is the similarity found in the presumed host specificity region of all the major lactococcal phage species: c2, 936, and all three P335 species (30). This is in accordance with the finding that some lactococcal strains are hosts for phages belonging to different species.

Outlook

Important aspects of lactococcal phage research have to some extent been limited by the difficulties in identifying sensitive lactococcal host strains. Transduction of a selectable marker is more sensitive than plaque assay, as shown for phages ΦLC3 and TP901-1, and will be a useful tool to identify host strains (9, 61). General transduction of chromosomal markers was only reported in 1962 (1), while plasmid transduction has been studied using phage T712 (43). Transduction of plasmids containing the *cos* site of the transducing phage has been reported for phages ΦLC3 and sk1 (9, 29). Development of general transducing systems for chromosomal markers would be a very helpful genetic tool for future research in *L. lactis*.

Functional analysis of the lactococcal phages by isolation and characterization of phage mutants is lagging behind the wealth of sequence data. The genetic tools exist, but have so far been used only in a few studies: Nonsense suppressors have been used in TP901-1 (93, 120) as well as gene disruptions (61). Clear-plaque mutants have been isolated in ΦLC3 (68) and TP901-1 (K. Hammer, unpublished data). Phages resistant to different naturally occurring abortive infection systems (*abi)* have been selected and analyzed (2, 40). An increased effort in the area of mutant isolation and characterization, combined with DNA-array analysis during infection, will give important information on structure-function analysis, including the gene regulatory mechanisms of the lactococcal phages and their interplay with the bacterial host. The existence of many different naturally occurring *abi* systems for lactococcal phages will greatly magnify the possibilities of insight into these important topics.

Note Added in Proof

The following findings of phage antireceptors involved in determination of host specificity have been published.

In the 936 phage species the Orf18 in sk1 (K.Dupont et al. 2004. Appl. Environ. Microbiol. 70:5818–5824). In the prolate c2 species the L10 protien is involved in addition to L15 (J. Rakonjac et al. 2005. J. Bacteriol. 187:3110–3121. In the P335 species ORF 49 of TP901-1 has been shown to form the lower baseplate disc of the phage tail and is suggested to function as the antireceptor (C. Vegge et al. 2005. J. Bacteriol. 187: in press).

Acknowledgments

We thank our colleagues S. McGrath, T. Klaenhammer, J. Kok, S. Leach, M. Lunde, I. Nes, I. B. Powell, D. van Sinderen, and B. Stuer-Lauridsen for communicating unpublished work to the authors. We are also grateful to D. van Sinderen and to our laboratory colleagues A. H. Johansen, M. Pedersen, and F. V. Vogensen for helpful comments during preparation of the manuscript. Financial support from the Danish Food research programs (FØTEK) through the Centre for Advanced Food Studies (LMC) and support from Director Ib Henriksens foundation is gratefully acknowledged.

References

1. Allen, L. K., W. E. Sandine, and P. R. Elliker. 1962. Transduction in *Streptococcus lactis*. J. Dairy Res. 30:351–357.
2. Allison, G. E., and T. R. Klaenhammer. 1998. Phage resistance mechanisms in lactic acid bacteria. Int. Dairy J. 8:207–226.
3. Anba, J., E. Bidnenko, A. Hillier, D. Ehrlich, and M. C. Chopin. 1995. Characterization of the lactococcal *abiD1* gene coding for phage abortive infection. J. Bacteriol. 177:3818–3823.
4. Arendt, E. K., C. Daly, G. F. Fitzgerald, and M. van de Guchte. 1994. Molecular characterization of lactococcal bacteriophage Tuc2009 and identification and analysis of genes encoding lysin, a putative holin, and two structural proteins. Appl. Environ. Microbiol. 60:1875–1883.
5. Beresford, T. P., L. J. H. Ward, and A. W. Jarvis. 1993. Temporally regulated transcriptional expression of the genomes of lactococcal bacteriophages c2 and sk1. Appl. Environ. Microbiol. 59:3708–3712.
6. Bidnenko, E., D. Ehrlich, and M. C. Chopin. 1995. Phage operon involved in sensitivity to the *Lactococcus lactis*

abortive infection mechanism AbiD1. J. Bacteriol. 177:3824–3829.

7. Bidnenko, E., S. D. Ehrlich, and M. C. Chopin. 1998. *Lactococcus lactis* phage operon coding for an endonuclease homologous to RuvC. Mol. Microbiol. 28:823–834.
8. Birkeland, N. K. 1994. Cloning, molecular characterization, and expression of the genes encoding the lytic functions of lactococcal bacteriophage ΦLC3: a dual lysis system of modular design. Can. J. Microbiol. 40:658–665.
9. Birkeland, N. K., and H. Holo. 1993. Transduction of a plasmid carrying the cohesive end region from *Lactococcus lactis* bacteriophage ΦLC3. Appl. Environ. Microbiol. 59:1966–1968.
10. Blatny, J. M., P. A. Risoen, D. Lillehaug, M. Lunde, and I. F. Nes. 2001. Analysis of a regulator involved in the genetic switch between lysis and lysogeny of the temperate *Lactococcus lactis* phage ΦLC3. Mol. Genet. Genomics 265:189–197.
11. Blatny, J. M., M. Ventura, E. M. Rosenhaven, P. A. Risoen, M. Lunde, H. Brussow, and I. F. Nes. 2003. Transcriptional analysis of the genetic elements involved in the lysogeny/lysis switch in the temperate lactococcal bacteriophage ΦLC3, and identification of the Cro-like proteins ORF76. Mol. Genet. Genomics 269:487–498.
12. Bolotin, A., P. Wincker, S. Mauger, O. Jaillon, K. Malarme, J. Weissenbach, S. D. Ehrlich, and A. Sorokin. 2001. The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403. Genome Res. 11:731–753.
13. Bouchard, J. D., and S. Moineau. 2000. Homologous recombination between a lactococcal bacteriophage and the chromosome of its host strain. Virology 270:65–75.
14. Boyce, J. D., B. E. Davidson, and A. J. Hillier. 1995. Identification of prophage genes expressed in lysogens of the *Lactococcus lactis* bacteriophage BK5-T. Appl. Environ. Microbiol. 61:4099–4104.
15. Braun, V., S. Hertwig, H. Neve, A. Geis, and M. Teuber. 1989. Taxonomic differentiation of bacteriophages of *Lactococcus lactis* by electron microscopy, DNA–DNA hybridization, and protein profiles. J. Gen. Microbiol. 135:2551–2560.
16. Breuner, A., L. Brøndsted, and K. Hammer. 1999. Novel organization of genes involved in prophage excision identified in the temperate lactococcal bacteriophage TP901-1. J. Bacteriol. 181:7291–7297.
17. Breuner, A., L. Brøndsted, and K. Hammer. 2001. Resolvase-like recombination performed by the TP901-1 integrase. Microbiology 147:2051–2063.
18. Brussow, H. 2001. Phages of dairy bacteria. Annu. Rev. Microbiol. 55:283–303.
19. Brussow, H., and F. Desiere. 2001. Comparative phage genomics and the evolution of *Siphoviridae*: insights from dairy phages. Mol. Microbiol. 39:213–222.
20. Bruttin, A., S. Foley, and H. Brüssow. 2002. DNA-binding activity of the *Streptococcus thermophilus* phage Sfi21 repressor. Virology 303:100–109.
21. Brøndsted, L., and K. Hammer. 1999. Use of the integration elements encoded by the temperate lactococcal bacteriophage TP901-1 to obtain chromosomal single-copy transcriptional fusions in *Lactococcus lactis*. Appl. Environ. Microbiol. 65:752–758.
22. Brøndsted, L., M. Pedersen, and K. Hammer. 2001. An activator of transcription regulates phage TP901-1 late gene expression. Appl. Environ. Microbiol. 67:5626–5633.
23. Brøndsted, L., S. Østergaard, M. Pedersen, K. Hammer, and F. K. Vogensen. 2001. Analysis of the complete DNA sequence of the temperate bacteriophage TP901-1: evolution, structure, and genome organization of lactococcal bacteriophages. Virology 283:93–109.
24. Callanan, M. J., P. W. O'Toole, M. W. Lubbers, and K. M. Polzin. 2001. Examination of lactococcal bacteriophage c2 DNA replication using two-dimensional agarose gel electrophoresis. Gene 278:101–106.
25. Casjens, S., G. F. Hatfull, and R. W. Hendrix. 1992. Evolution of dsDNA tailed-bacteriophage genomes. Semin. Virol. 3:383–397.
26. Chandry, P. S., B. E. Davidson, and A. J. Hillier. 1994. Temporal transcription map of the *Lactococcus lactis* bacteriophage sk1. Microbiology 140:2251–2261.
27. Chandry, P. S., S. C. Moore, J. D. Boyce, B. E. Davidson, and A. J. Hillier. 1997. Analysis of the DNA sequence, gene expression, origin of replication and modular structure of the *Lactococcus lactis* lytic bacteriophage sk1. Mol. Microbiol. 26:49–64.
28. Chandry, P. S., S. C. Moore, B. E. Davidson, and A. J. Hillier. 1994. Analysis of the *cos* region of the *Lactococcus lactis* bacteriophage sk1. Gene 138:123–126.
29. Chandry, P. S., S. C. Moore, B. E. Davidson, and A. J. Hillier. 2002. Transduction of contatemeric plasmids containing the *cos* site of *Lactococcus lactis* bacteriophage sk1. FEMS Microbiol. Lett. 216:85–90.
30. Chopin, A., A. Bolotin, A. Sorokin, S. D. Ehrlich, and M. Chopin. 2001. Analysis of six prophages in *Lactococcus lactis* IL1403: different genetic structure of temperate and virulent phage populations. Nucleic Acids Res. 29:644–651.
31. Christiansen, B., L. Brøndsted, F. K. Vogensen, and K. Hammer. 1996. A resolvase-like protein is required for the site-specific integration of the temperate lactococcal bacteriophage TP901-1. J. Bacteriol. 178:5164–5173.
32. Christiansen, B., M. G. Johnsen, E. Stenby, F. K. Vogensen, and K. Hammer. 1994. Characterization of the lactococcal temperate phage TP901-1 and its site-specific integration. J. Bacteriol. 176:1069–1076.
33. Chung, D. K., J. H. Kim, and C. A. Batt. 1991. Cloning and nucleotide sequence of the major capsid protein of *Lactococcus lactis* ssp. *cremoris* bacteriophage F4-1. Gene 101:121–125.
34. Crutz-Le Coq, A. M., B. Cesselin, J. Commissaire, and J. Anba. 2002. Sequence analysis of the lactococcal bacteriophage bIL170: insights into structural proteins and HNH endonucleases in dairy phages. Microbiology 148:985–1001.
35. Desiere, F., C. Mahanivong, A. J. Hillier, P. S. Chandry, B. E. Davidson, and H. Brussow. 2001. Comparative genomics of lactococcal phages: insight from the complete genome sequence of *Lactococcus lactis* phage BK5-T. Virology 283:240–252.
36. Desiere, F., W. M. McShan, D. van Sinderen, J. J. Ferretti, and H. Brussow. 2001. Comparative genomics reveals close genetic relationships between phages from dairy bacteria

and pathogenic streptococci: evolutionary implications for prophage–host interactions. Virology 288:325–341.

37. Durmaz, E., and T. R. Klaenhammer. 2000. Genetic analysis of chromosomal regions of *Lactococcus lactis* acquired by recombinant lytic phages. Appl. Environ. Microbiol. 66:895–903.
38. Durmaz, E., S. M. Madsen, H. Israelsen, and T. R. Klaenhammer. 2002. *Lactococcus lactis* lytic bacteriophages of the P335 group are inhibited by overexpression of a truncated CI repressor. J. Bacteriol. 184:6532–6544.
39. Duwat, P., Hammer, K., Bolotin, A., and Gruss, A. 1999. Genetics of lactococci, pp. 295–306. In V. Fischetti, R. Novick, J. Ferretti, D. Portnoy, and J. Rood (eds.), Gram-positive Pathogens. ASM Press, Washington, D.C.
40. Forde, A., and G. F. Fitzgerald . 1999. Bacteriophage defence systems in lactic acid bacteria. Antonie van Leeuwenhoek 76:89–113.
41. Fox, P. F. 1989. Cheese: an overview, pp. 1–36. *In* P. F. Fox (ed.) Cheese: Chemistry, Physics and Microbiology. Chapman and Hall, London.
42. Garbutt, K. C., J. Kraus, and B. L. Geller. 1997. Bacteriophage resistance in *Lactococcus lactis* engineered by replacement of a gene for a bacteriophage receptor. J. Dairy Sci. 80:1512–1519.
43. Gasson, M., and Fitzgerald, G. F. 1993. Gene transfer systems and transposition, pp. 1–51. *In* M. Gas, and W. de Vas (eds.) Genetics and Biotechnology of Lactic Acid Bacteria. Chapman and Hall, London.
44. Geller, B. L., R. G. Ivey, J. E. Trempy, and B. Hettinger-Smith. 1993. Cloning of a chromosomal gene required for phage infection of *Lactococcus lactis* subsp. *lactis* C2. J. Bacteriol. 175:5510–5519.
45. Godon, J. J., C. Delorme, S. D. Ehrlich, and P. Renault. 1992. Divergence of genomic sequences between *Lactococcus lactis* subsp. *lactis* and *Lactococcus lactis* subsp. *cremoris*. Appl. Environ. Microbiol. 58:4045–4047.
46. Hatfull, G. F., and G. J. Sarkis. 1993. DNA sequence, structure and gene expression of mycobacteriophage L5: a phage system for mycobacterial genetics. Mol. Microbiol. 7:395–405.
47. Heap, H. A., G. K. Y. Limsowtin, and R. C. Lawrence. 1978. Contribution of *Streptococcus lactis* strains in raw milk to phage infection in cheese factories. N. Z. J. Dairy Sci. Technol. 13:16–22.
48. Hendrix, R. W., M. C. Smith, R. N. Burns, M. E. Ford, and G. F. Hatfull. 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. Proc. Natl. Acad. Sci. USA 96:2192–2197.
49. Hill, C., L. A. Miller, and T. R. Klaenhammer. 1990. Cloning, expression, and sequence determination of a bacteriophage fragment encoding bacteriophage resistance in *Lactococcus lactis*. J. Bacteriol. 172:6419–6426.
50. Hill, C., L. A. Miller, and T. R. Klaenhammer. 1991. In vivo genetic exchange of a functional domain from a type II A methylase between lactococcal plasmid pTR2030 and a virulent bacteriophage. J. Bacteriol. 173:4363–4370.
51. Huggins, A. R., and W. E. Sandine. 1977. Incidence and properties of temperate bacteriophages induced from lactic streptococci. Appl. Environ. Microbiol. 33:184–191.
52. Jarvis, A. W. 1984. Differentiation of lactic streptococcal phages into phage species by DNA–DNA homology. Appl. Environ. Microbiol. 47:343–349.
53. Jarvis, A. W. 1989. Bacteriophages of lactic acid bacteria. J. Dairy Sci. 72:3406–3428.
54. Jarvis, A.W. 1995. Relationships by DNA homology between lactococcal phages 7-9, P335 and New Zealand industrial lactococcal phages. Int. Dairy J. 5:355–366.
55. Jarvis, A.W., G. F. Fitzgerald, M. Mata, A. Mercenier, H. Neve, I. B. Powell, C. Ronda, M. Saxelin, and M. Teuber. 1991. Species and type phages of lactococcal bacteriophages. Intervirology 32:2–9.
56. Johansen, A. H., L. Brøndsted, and K. Hammer. 2003. Identification of operator sites of the CI repressor of phage TP901-1: evolutionary link to other phages. Virology 311:144–156.
57. Johnsen, M. G. 1995. Virion proteins and genetics of lactococcal temperate phage TP901–1. PhD thesis, The Royal Veterinary and Agricultural University, Denmark.
58. Johnsen, M. G., K. F. Appel, P. L. Madsen, F. K. Vogensen, K. Hammer, and J. Arnau. 1996. A genomic region of lactococcal temperate bacteriophage TP901-1 encoding major virion proteins. Virology 218:306–315.
59. Johnsen, M. G., H. Neve, F. K. Vogensen, and K. Hammer. 1995. Virion positions and relationships of lactococcal temperate bacteriophage TP901-1 proteins. Virology 212:595–606.
60. Josephsen, J., N. Andersen, H. Behrndt, E. Brandsborg, G. Christiansen, M. B. Hansen, S. Hansen, E. Waagner Nielsen, and F. K. Vogensen. 1994. An ecological study of lytic bacteriophages of *Lactococcus lactis* subsp. *cremoris* isolated in a cheese plant over a five year period. Int. Dairy J. 4:123–141.
61. Koch, B., B. Christiansen, T. Evison, F. K. Vogensen, and K. Hammer. 1997. Construction of specific erythromycin resistance mutations in the temperate lactococcal bacteriophage TP901-1 and their use in studies of phage biology. Appl. Environ. Microbiol. 63:2439–2441.
62. Kok, J., O. Kuipers., A. Jong, D. van Sinderen, G. Fitzgerald, U. Wegmann, C. Shearman, and M. Gasson. 2001. The genome sequencing project of *Lactococcus lactis* MG1363. Book of abstracts, EuroLAB conference, University College, Cork, Ireland.
63. Kraus, J., and B. L. Geller. 1998. Membrane receptor for prolate phages is not required for infection of *Lactococcus lactis* by small or large isometric phages. J. Dairy Sci. 81:2329–2335.
64. Kuhstoss, S., and R. N. Rao. 1991. Analysis of the integration function of the streptomycete bacteriophage ΦC31. J. Mol. Biol. 222:897–908.
65. Labrie, S., and S. Moineau. 2002. Complete genome sequence of bacteriophage ul36: demonstration of phage heterogeneity within the P335 quasi-species of lactococcal phages. Virology 296:308–320.
66. Leach, S., D. van Sinderen, J. Kok, G. F. Fitzgerald, G. Venema, and A. Nauta. 1999. Molecular analysis of the genetic switch of Tuc2009. Book of abstracts, EuroLAB conference, University College, Cork, Ireland.
67. Levine, M. E., R. W. Hendrix, and S. R. Casjens. 1993. A programmed translational frameshift is required for the

synthesis of a bacteriophage-lambda tail assembly protein. J. Mol. Biol. 234:124–139.

68. Lillehaug, D., and N. K. Birkeland. 1993. Characterization of genetic elements required for site-specific integration of the temperate lactococcal bacteriophage ΦLC3 and construction of integration-negative ΦLC3 mutants. J. Bacteriol. 175:1745–1755.
69. Lillehaug, D., B. Lindqvist, and N. K. Birkeland. 1991. Characterization of ΦLC3, a *Lactococcus lactis* subsp. *cremoris* temperate bacteriophage with cohesive single-stranded DNA ends. Appl. Environ. Microbiol. 57:3206–3211.
70. Lillehaug, D., I. F. Nes, and N. K. Birkeland. 1997. A highly efficient and stable system for site-specific integration of genes and plasmids into the phage ΦLC3 attachment site (*att*B) of the *Lactococcus lactis* chromosome. Gene 188:129–136.
71. Lubbers, M.W., N. R.Waterfield,T. P. Beresford, R.W. Le Page, and A. W. Jarvis. 1995. Sequencing and analysis of the prolate-headed lactococcal bacteriophage c2 genome and identification of the structural genes. Appl. Environ. Microbiol. 61:4348–4356.
72. Lucchini, S., F. Desiere, and H. Brussow. 1999. Similarly organized lysogeny modules in temperate *Siphoviridae* from low GC content Gram-positive bacteria. Virology 263:427–435.
73. Lunde, M., J. M. Blatny, F. Kaper, I. F. Nes, and D. Lillehaug. 2000. The life cycles of the temperate lactococcal bacteriophage ΦLC3 monitored by a quantitative PCR method. FEMS Microbiol. Lett. 192:119–124.
74. Lunde, M., J. M. Blatny, D. Lillehaug, A. H. Aastveit, and I. F. Nes. 2003. Use of real-time quantitative PCR for the analysis of ΦLC3 prophage stability in lactococci. Appl. Environ. Microbiol. 69:41–48.
75. Madsen, P. L., and K. Hammer. 1998. Temporal transcription of the lactococcal temperate phage TP901-1 and DNA sequence of the early promoter region. Microbiology 144:2203–2215.
76. Madsen, P. L., A. H. Johansen, K. Hammer, and L. Brøndsted. 1999. The genetic switch regulating activity of early promoters of the temperate lactococcal bacteriophage TP901-1. J. Bacteriol. 181:7430–7438.
77. Madsen, S. M., D. Mills, G. Djordjevic, H. Israelsen, and T. R. Klaenhammer. 2001. Analysis of the genetic switch and replication region of a P335-type bacteriophage with an obligate lytic lifestyle on *Lactococcus lactis*. Appl. Environ. Microbiol. 67:1128–1139.
78. Mahanivong, C., J. D. Boyce, B. E. Davidson, and A. J. Hillier. 2001. Sequence analysis and molecular characterization of the *Lactococcus lactis* temperate bacteriophage BK5-T. Appl. Environ. Microbiol. 67:3564–3576.
79. Matsuura, M., T. Noguchi, D. Yamaguchi, T. Aida, M. Asayama, H. Takahashi, and M. Shirai. 1996. The *sre* gene (ORF469) encodes a site-specific recombinase responsible for integration of the R4 phage genome. J. Bacteriol. 178:3374–3376.
80. McGrath, S., G. F. Fitzgerald, and D. van Sinderen. 2001. Improvement and optimization of two engineered phage resistance mechanisms in *Lactococcus lactis*. Appl. Environ. Microbiol. 67:608–616.
81. McGrath, S., G. F. Fitzgerald, and D. van Sinderen. 2002. Identification and characterisation of phage-resistance genes in temperate lactococcal bacteriophages. Mol. Microbiol. 43:509–520.
82. McGrath, S., J. F. Seegers, G. F. Fitzgerald, and D. van Sinderen. 1999. Molecular characterization of a phage-encoded resistance system in *Lactococcus lactis*. Appl. Environ. Microbiol. 65:1891–1899.
83. Moineau, S., S. A. Walker, E. R. Vedamuthu, and P. A. Vandenbergh. 1995. Cloning and sequencing of *Lla*II restriction/modification genes from *Lactococcus lactis* and relatedness of this system to the *Streptococcus pneumoniae Dpn*II system. Appl. Environ. Microbiol. 61:2193–2202.
84. Monteville, M. R., B. Ardestani, and B. L. Geller. 1994. Lactococcal bacteriophages require a host cell wall carbohydrate and a plasma membrane protein for adsorption and ejection of DNA. Appl. Environ. Microbiol. 60:3204–3211.
85. Nauta, A. 1997. Molecular characterization and exploitation of the temperate *Lactococcus lactis* bacteriophage r1t. PhD thesis. Department of Genetics, University of Groningen, The Netherlands.
86. Nauta, A., B. B. van den, H. Karsens, G. Venema, and J. Kok. 1997. Design of thermolabile bacteriophage repressor mutants by comparative molecular modelling. Nat. Biotechnol. 15:980–983.
87. Nauta, A., D. van Sinderen, H. Karsens, E. Smit, G. Venema, and J. Kok. 1996. Inducible gene expression mediated by a repressor-operator system isolated from *Lactococcus lactis* bacteriophage r1t. Mol. Microbiol. 19:1331–1341.
88. Neve, H., K. I. Zenz, F. Desiere, A. Koch, K. J. Heller, and H. Brussow. 1998. Comparison of the lysogeny modules from the temperate *Streptococcus thermophilus* bacteriophages TP-J34 and Sfi21: implications for the modular theory of phage evolution. Virology 241:61–72.
89. O'Sullivan, D. J., S. A. Walker, S. G. West, and T. R. Klaenhammer. 1996. Development of an expression strategy using a lytic phage to trigger explosive plasmid amplification and gene expression. Biotechnology (N.Y.) 14:82–87.
90. Parreira, R., S. D. Ehrlich, and M. C. Chopin. 1996. Dramatic decay of phage transcripts in lactococcal cells carrying the abortive infection determinant AbiB. Mol. Microbiol. 19:221–230.
91. Parreira, R., R. Valyasevi, A. L. Lerayer, S. D. Ehrlich, and M. C. Chopin. 1996. Gene organization and transcription of a late-expressed region of a *Lactococcus lactis* phage. J. Bacteriol. 178:6158–6165.
92. Pedersen, M. 2003. Regulation of the late transcription in the lactococcal phage TP901–1. PhD thesis, Technical University of Denmark, Denmark.
93. Pedersen, M., S. Østergaard, J. Bresciani, and F. K. Vogensen. 2000. Mutational analysis of two structural genes of the temperate lactococcal bacteriophage TP901-1 involved in tail length determination and baseplate assembly. Virology 276:315–328.
94. Petersen, A., J. Josephsen, and M. G. Johnsen. 1999. TPW22, a lactococcal temperate phage with a site-specific integrase closely related to *Streptococcus thermophilus* phage integrases. J. Bacteriol. 181:7034–7042.

95. Pillidge, C. J., and A. W. Jarvis. 1988. DNA restriction maps and classification of lactococcal bacteriophages c2 and sk1. N. Z. J. Dairy Sci. Technol. 23:411–416.
96. Popa, M. P., T. A. McKelvey, J. Hempel, and R. W. Hendrix. 1991. Bacteriophage HK97 structure: wholesale covalent cross-linking between the major head shell subunits. J. Virol. 65:3227–3237.
97. Prévots, F., M. Mata, and P. Ritzenthaler. 1990. Taxanomic differentiation of 101 lactococcal bacteriophages and characterization of bacteriophages with unusual large genomes. Appl. Environ. Microbiol. 56:2180–2185.
98. Reiter, B. 1949. Lysogenic strains of lactic streptococci. Nature 164:667–668.
99. Rince, A., M. Tangney, and G. F. Fitzgerald. 2000. Identification of a DNA region from lactococcal phage sk1 protecting phage 712 from the abortive infection mechanism AbiF. FEMS Microbiol. Lett. 182:185–191.
100. Sandine, W. E., P. C. Radich, and P. R. Elliker. 1972. Ecology of the lactic streptococci. A review. J. Food Technol. 35:176–184.
101. Schleifer, K. H., J. Kraus, C. Dvorak, R. Kilpper-Bälz, M. D. Collins, and W. Fischer. 1985. Transfer of *Streptococcus lactis* and related streptococci to the genus *Lactococcus* gen. nov. Syst. Appl. Microbiol. 6:183–195.
102. Schouler, C., S. D. Ehrlich, and M. C. Chopin. 1994. Sequence and organization of the lactococcal prolate-headed bIL67 phage genome. Microbiology 140: 3061–3069.
103. Sharples, G. J., E. L. Bolt, and R. G. Lloyd. 2002. RusA proteins from the extreme thermophile *Aquifex aeeolicus* and lactococcal phage r1t resolve Holliday junctions. Mol. Microbiol. 44:549–559.
104. Sheehan, M. M., E. Stanley, G. F. Fitzgerald, and D van Sinderen. 1999. Identification and characterization of a lysis module present in a large proportion of bacteriophages infecting *Streptococcus thermophilus*. Appl. Environ. Microbiol. 65:569–577.
105. Stark, W. M., M. R. Boocock, and D. J. Sherrat. 1989. Site-specific recombination by Tn3 resolvase. Trends Genet. 5:304–309.
106. Steidler, L., W. Hans, L. Schotte, S. Neirynck, F. Obermeier, W. Falk, W. Fiers, and E. Remaut. 2000. Treatment of murine colitis by *Lactococcus lactis* secreting interleukin-10. Science 289:1352–1355.
107. Stoll, S. M., D. S. Ginsburg, and M. P. Calos. 2002. Phage TP901-1 site-specific integrase functions in human cells. J. Bacteriol. 184: 3657–3663.
108. Stuer-Lauridsen, B., T. Janzen, J. Schnabl, and E. Johansen. 2003. Identification of the host determinant of two-prolate-headed phages infecting *Lactococcus lactis*. Virology 309:10–17.
109. Thorpe, H. M., and M. C. Smith. 1998. In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. Proc. Natl. Acad. Sci. USA 95:5505–5510.
110. van de Guchte, M., C. Daly, G. F. Fitzgerald, and E. K. Arendt. 1994. Identification of *int* and *att*P on the genome of lactococcal bacteriophage Tuc2009 and their use for site-specific plasmid integration in the chromosome of Tuc2009-resistant *Lactococcus lactis* MG1363. Appl. Environ. Microbiol. 60:2324–2329.
111. van Sinderen, D., H. Karsens, J. Kok, P. Terpstra, M. H. Ruiters, G. Venema, and A. Nauta. 1996. Sequence analysis and molecular characterization of the temperate lactococcal bacteriophage r1t. Mol. Microbiol. 19:1343–1355.
112. Venema, G., J. Kok, and D. van Sinderen. 1999. From DNA sequence to application: possibilities and complications. Antonie van Leeuwenhoek 76:3–23.
113. Walker, S. A., C. S. Dombroski, and T. R. Klaenhammer. 1998. Common elements regulating gene expression in temperate and lytic bacteriophages of *Lactococcus* species. Appl. Environ. Microbiol. 64:1147–1152.
114. Walker, S. A., and T. R. Klaenhammer. 1998. Molecular characterization of a phage-inducible middle promoter and its transcriptional activator from the lactococcal bacteriophage Φ31. J. Bacteriol. 180:921–931.
115. Ward, L. J. H., T. P. J. Beresford, M. W. Lubbers, B. D. W. Jarvis, and A. W. Jarvis. 1993. Sequence-analysis of the lysin gene region of the prolate lactococcal bacteriophage c2. Can. J. Microbiol. 39:767–774.
116. Waterfield, N. R., M. W. Lubbers, K. M. Polzin, R. W. Le Page, and A. W. Jarvis. 1996. An origin of DNA replication from *Lactococcus lactis* bacteriophage c2. Appl. Environ. Microbiol. 62:1452–1453.
117. Whitehead, H. R., and G. A. Cox. 1935. The occurrence of bacteriophages in culture of lactic streptococci. N. Z. J. Dairy Sci. Technol. 16:319–320.
118. Williams, A. H., J. L. Fryers, and M. D. Collins. 1990. *Lactococcus piscium* sp. nov. a new *Lactococcus* species from salmonid fish. FEMS Microbiol. Lett. 68:109–114.
119. Zuniga, M., B. Franke-Fayard, G. Venema, J. Kok, and A. Nauta. 2002. Characterization of the putative replisome organizer of the lactococcal bacteriophage r1t. J. Virol. 76:10234–10244.
120. Østergaard, S., L. Brøndsted, and F. K. Vogensen. 2001. Identification of a replication protein and repeats essential for DNA replication of the temperate lactococcal bacteriophage TP901-1. Appl. Environ. Microbiol. 67:774–781.

37

The *Listeria* Bacteriophages

MARTIN J. LOESSNER
RICHARD CALENDAR

Bacterial viruses specific for the genus *Listeria* were discovered almost 60 years ago (57), and were early reported for their usefulness in phage typing (61) of isolates of the pathogen *Listeria monocytogenes* (65). In the following years, phage typing of *Listeria* isolates has proven to be a very useful method, and led to the isolation of more than 400 phages for *L. monocytogenes*, *L. ivanovii*, and the nonpathogenic species *L. innocua*, *L. seeligeri*, and *L. welshimeri* (5, 9, 16, 19, 22, 25, 26, 31–33, 48–51, 53, 55, 63). To date, no phages infecting organisms of the species *L. grayi* have been found. This chapter briefly summarizes our present knowledge on *Listeria* phages, and gives an overview on their general and particular properties, with respect to both the basic science and the various practical applications.

Ultrastructure, Composition, and Taxonomy of *Listeria* Phages

Electron microscopical examinations of more than 120 *Listeria* phages (2, 5, 7, 10, 33, 49, 51, 61, 71) revealed a relatively limited diversity (figure 37-1). Most phages belong to the morphotype group B1 of the *Siphoviridae* family (isometric capsid and long, noncontractile tail), in the order *Caudovirales* (1) (see chapter 2 for an overview of phage classification). These Siphoviruses are divided into five recognized species and one proposed species (table 37-1) based upon differences in tail length (3). The remaining phages were classified as *Myoviridae* of the morphotype A1 (isometric capsid and long, contractile tail). Two species were established, based on different particle dimensions and mode of sheath contraction. While members of the species A511 resemble a more commonly found phage morphotype, phages of species 4211 (such as 01761 depicted in figure 37-1) are unique and feature a rather unusual mode of sheath contraction in which the tail seems to contract toward the baseplate, thereby exposing the inner tail tube starting from the capsid.

Caudovirales generally have double-stranded DNA as genetic material. Restriction endonuclease analysis allowed discrimination of individual *Listeria* phages and calculation of genome sizes, which range from 36 to more than 100 kb (table 37-1). The G-C content lies between 35 and 41 mol % (36, 38, 54, 70; unpublished data). A significant correlation between ultrastructure and overall DNA homology was found, which supports the existence of at least five DNA homology groups among the phages investigated (38, 54). Structural proteins of more than 40 phages were analyzed by electrophoretic methods (33, 38, 70, 71), and, very recently, mass spectrometry (70). Several studies indicate that the major capsid proteins are proteolytically processed during maturation of the head, while from the tail sheath proteins only the N-terminal methionine is absent (36, 41, 70). The PSA virion employs a particularly interesting mechanism for synthesis of essential components, involving a translational +1 frameshifting at the 5′ ends of the major structural protein genes for the capsid (*cps*) and tail (*tsh*) (70). Comparison of protein profiles permits differentiation of phages and establishment of similarity clusters, and generally corresponds well with ultrastructure and DNA hybridization patterns.

Host Ranges and Phage Receptors

All *Listeria* phages are strictly genus-specific. The temperate ones only recognize host bacteria of individual but specific serovar groups, while the virulent A511-like phages can attack strains of all species and serovars (4, 5, 31, 32, 50, 52). The *Listeria* O-antigens are largely determined by the variable structure and sugar substitution of polyribitolphosphate cell wall teichoic acids (17). It has been demonstrated by biochemical and genetical approaches that the teichoic acid substituents, *N*-acetylglucosamine and rhamnose, are major determinants of phage adsorption in serovar 1/2 strains, while *N*-acetylglucosamine and

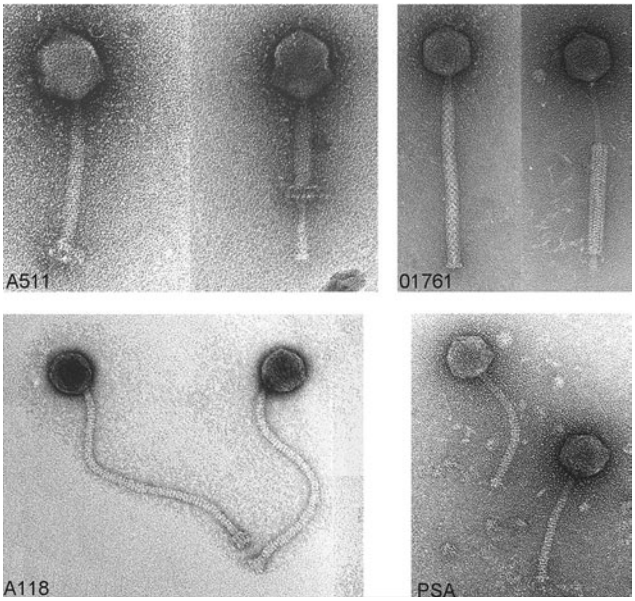


Figure 37-1 Electron micrographs of different *Listeria* bacteriophages. The phages shown belong to different *morphotypes* and species. A511 and 01761 are *Myoviridae* with contractile tails, while A118 and PSA are *Siphoviridae* with noncontractile, flexible tails (see table 37-1 and text).

galactose are important in serovar 4 strains (11, 62, 68). In contrast, teichoic acids are apparently not involved in binding of the polyvalent A511 phage. It is assumed that the peptidoglycan itself represents its receptor, possibly in conjunction with other, serovar-independent carbohydrates (68).

Virus Multiplication and Host Cell Lysis

Listeria phages seem to be well adapted to their host bacteria. Most phages can complete lytic cycles at 10–37°C, but some are more temperature-sensitive and only multiply at 25°C or below (25). Growth curves have been recorded for a few phages. At 30°C, the latent period of the lytic cycle of temperate phages infecting *L. monocytogenes* is between 60 and 70 minutes, followed by a rise phase of 40 to 65 minutes. An average of 25 progeny virions are released from lysed cells. In *L. ivanovii* as host, latent phases up to 115 minutes were observed, resulting in a burst size of up to 40 particles. The virulent nature of A511 agrees well with its shorter latent phase of 55 minutes and comparatively large burst size of 40 virions (unpublished data). (For an overview of phage infection-timing and burst-size characters, see chapter 5.)

Lysis of infected cells occurs through the combined action of a holin (Hol) and an endolysin (Ply), encoded by two immediately adjacent genes at the distal end of the late gene regions. In phage A511, the holin has not been identified. In phages A118, A500, and ϕ2438, Ply (the endolysin) was found to represent a new class of enzyme, an L-alanoyl-*D*-glutamate endopeptidase, whereas phages A511 and PSA feature (different) *N*-acetylmuramoyl-L-alanine amidases (44, 70, 73). An interesting aspect of these enzymes is their modular composition and their unusual substrate specificity, which is mediated by individual C-terminal domains recognizing unique cell wall carbohydrates (37). These cell wall binding domains (CBD) have distinct binding abilities: The CBD of phage Ply500 endolysin recognizes cell surfaces belonging to serovars 4, 5, and 6, and binds to a receptor evenly distributed in the wall. In contrast,

Table 37-1 Present Status of *Listeria* Phage Taxonomy, and Main Virion Characteristics

Family	Species	Other relevant members[a]	Approximate[b] virion size (in nm) (head diameter[c]/tail length[d])	Approximate genome sizes (in kb)	References
Myoviridae	A511		88/200	130–140	31, 38, 44, 71 (unpublished)
	4211	B054 01761	62–66/230–270	41–44	33, 51, 71
Siphoviridae	P35[e]		56–60/110	36	25, (unpublished)
	2389	PSA	58–62/170–180	38	2, 33, 70
	H387		58–62/190–200	36–40	49
	2685	B025	58–62/230–260	37–41	33, 54, 71
	2671	A118 A500 A006	58–62/270–310	38–42	25, 33, 36, 54, 71, (unpublished)

[a] Phages investigated and sequenced in our laboratories.

[b] Different phages within the species, different isolates, and different staining methods yield variable results.

[c] Measured from apex to apex.

[d] Measured including base plates.

[e] Proposed new species.

the CBD of phage Ply118 binds only to serovar 1/2 and 3 cells, with preference for septal regions and the poles. These distinctions indicate fundamental differences in carbohydrate composition among the cell walls of different strains of bacteria as well as cell wall differences along the contours of individual bacteria.

Immediately preceding cell wall hydrolysis, Hol is proposed to form lesions in the cytoplasmic membrane, enabling access of Ply to the murein (reviewed in chapter 10). Although the primary sequences of Hol118 and Hol500 (the Hol proteins from phages A118 and A500, respectively) are completely different from the prototype phage λ S protein, the overall structures of these class II holins are conserved (44). However, timing and regulation of Hol118 function is clearly different from the S paradigm (67). Although it features a dual-start motif, recent findings have demonstrated that Hol118-mediated regulation of lysis timing represents a novelty since it occurs via a cotranscribed intragenic inhibitor lacking the first transmembrane domain and thereby interfering with pore formation (66).

Temperate *Listeria* Phages

Genomics

The first *Listeria* phage completely sequenced and analyzed in detail was A118, a temperate bacteriophage specific for *L. monocytogenes* serovar 1/2 strains (36). Its genome is a circularly permuted collection of terminally redundant dsDNA molecules of an average length of 43.3 kb, which indicates 6% redundancy of the unit size of 40,834 bp. This has been confirmed by partial denaturation maps and electron microscopy, also showing that the right end of DNA is attached to the phage tail and that the roughly 10-mer concatemers are sequentially packaged left to right. The A118 genome contains 72 ORFs, organized in three major, life-cycle-specific gene clusters. The genes required for lytic development show an opposite orientation and arrangement compared with the lysogeny control region. A function could be assigned to 26 genes, while the remaining 46 have no known counterparts.

The sequence of the prophage of *L. monocytogenes* serovar4b strain ScottA, PSA, was recently completed (70). In contrast to A118, PSA features 10-nucleotide 3′-overhanging cohesive ends and packages exactly one unit genome of 37,618 bp. Although its 55 open reading frames are mostly unrelated to phage A118's (figure 37-2), their overall life-style-specific gene organization is relatively similar, except for the presence of some unique genes such as a primase and a helicase in PSA, and a recombinase in A118, each of which are required for the different mechanisms of genome replication.

Proteome analysis of PSA revealed an unusual form of translational frameshifting which yields different-length forms of the major structural proteins of the capsid and tail, respectively (70). The proteins feature identical N-termini but different C-termini as a result of programmed +1 translational frameshifting. Frameshifting appears to be initiated by a slippery nucleotide sequence with overlapping proline codons near the 3′ ends of both genes. This apparently redirects the ribosomes into the +1 frames. Different *cis*-acting factors (a shifty stop and a pseudoknot) are also present. This phage PSA attribute is the first case of +1 frameshift among double-standed DNA phages, and also is the prototype of a virus featuring a 3′ pseudoknot to stimulate ribosomal frameshifts.

Attachment Sites and Integration

So far, two different classes of temperate phages in *L. monocytogenes* have been shown to integrate their DNA into the host chromosome. Phage A118 is an example of the first class (36). Sequence comparisons indicate that the A118 integrase enzyme is a serine recombinase related to Tn10 resolvase and Hin invertase. The A118 bacterial attachment site, *attB*, lies within an open reading frame closely related to *comK* of *Bacillus subtilis*. This gene encodes a transcriptional activator for various factors involved in competence for DNA uptake. Since *L. monocytogenes* is not easily transformable, the role of its *comK* gene is not immediately obvious. Integration of the A118 genome into *comK* changes most of the coding sequence, and no phage sequence reconstitutes this reading frame. The A118 phage and bacterial attachment sites display only 3 base pairs of homology near the point of crossover, as is common for serine recombinases. In contrast, phage PSA integrates into the tRNA_{Arg} gene (29). PSA's phage attachment site contains identity to the 15 nucleotides at the 3′ end of tRNA_{Arg} plus two downstream nucleotides. After integration by PSA, the sequence of tRNA_{Arg} is regenerated by prophage nucleotides. PSA's integrase is different from A118 Int, and is homologous to *Escherichia coli* XerD, a tyrosine recombinase that resolves dimeric circles of *E. coli* DNA and plasmids of the ColE1 family (60).

Prophages and Lysogeny

Lysogeny is widespread among strains of the genus *Listeria*; the percentage of strains producing infective phage has been estimated to range from 6% to 37%, depending on the species (50). Prophages are readily inducible using UV light or mitomycin C (34). Lysogenization can easily be provoked by a high multiplicity of infection, and lysogens are generally resistant to superinfection by the same or related phages but not by phages of different immunity groups (35). Many of the commonly used laboratory strains (e.g., 10403S, EGDe, LO28, and others) carry an integrated functional or cryptic prophage at the *attB* within *comK*. However, *comK* and $tRNA_{arg}$ are only two of several

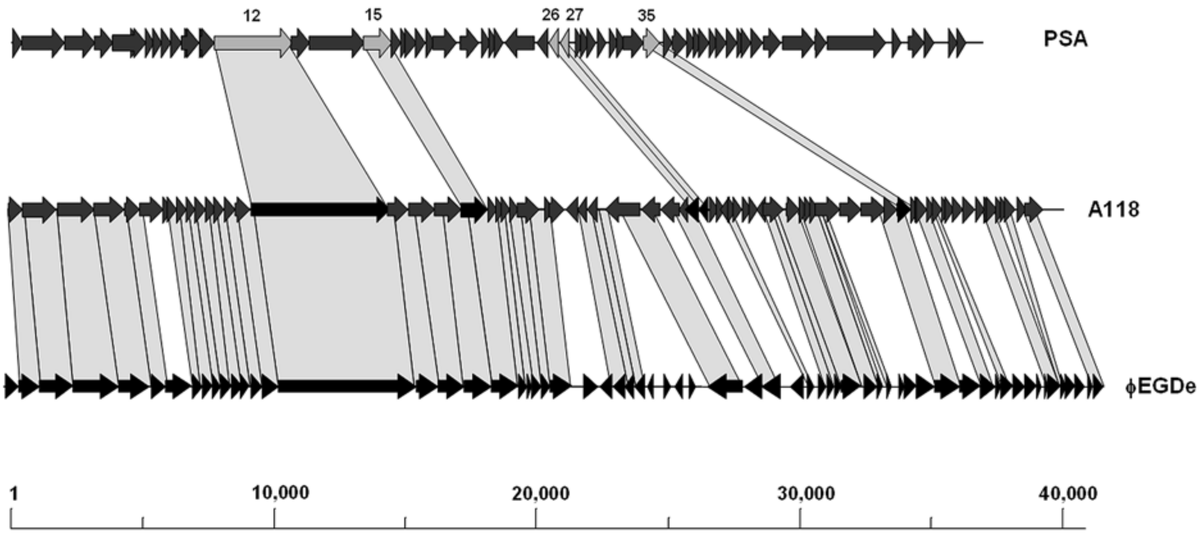


Figure 37-2 Alignment of the genetic maps from the temperate *Listeria monocytogenes* phages. Shown are phages PSA (70), A118 (36), and the cryptic prophage ϕEGDe (20). Open reading frames are shown as black arrows, pointing in the direction of transcription. The maps start with the terminase genes of the ''late'' region (left), and end with distal genes of the ''early'' gene region, involved in DNA replication and recombination. Lysogeny control regions (mostly leftward-pointing arrows) are located at approximate coordinates 20–30 kb on the ruler at the bottom of the figure. Genes encoding proteins of significant amino acid sequence similarity are linked by gray shading. For PSA, the few genes whose products feature significant similraity to A118 are indicated by numbers corresponding to the individual open reading frames (70).

existing *attB* sites in the *Listeria* genomes, since multiple lysogens can be created by subsequent challenge with different phages (35), and polylysogenic strains are frequently observed (see chapter 7 for a discussion of phage integration).

Despite the fact that most if not all strains carry functional or cryptic prophages, the potential influence of lysogeny on the host phenotype is unknown. No phenotype has been associated with *comK* inactivation by insertion of A118-like phage or by insertion of an integration vector (described below). No obvious association was yet observed between phage carrier state and *Listeria* phenotype, especially with regard to pathogenicity. Nevertheless, temperate phage may carry genes similar to host factors involved in bacteria–host interaction (36). Resistance to phage was sometimes found to be the result of changes in cell wall composition (62, 68), which could also be linked to decreased sensitivity to quaternary ammonium compounds (46). Moreover, the presence of phage-encoded methyltransferases such as M.*Lmo*A118I (8, 36) affects distribution of genetic material and may therefore influence the phenotype of lysogens.

Cryptic Prophages

Up to 71% (27) of all *Listeria* cultures produce substances inhibitory to other *Listeria* strains but not to other bacteria. These substances were termed "monocins" (7, 12, 22, 47, 61, 69). In monocin preparations, particles that resemble phage tails or polyheads could be observed by electron microscopy (7, 73). It was later shown by genetic methods that these particles indeed result from incomplete, cryptic prophages and that their lethal effect is due to the presence of intact lysis genes (73). Similar to phage, monocins display a killing-from-without effect which to some degree is serovar correlated (72). A cryptic phage related to A118 was identified in the chromosome of *L. monocytogenes* EGDe (see figure 37-1), and five phage-like elements are present in the chromosome of *L. innocua* CLIP11262 (20).

Transducing Phages

Many of the temperate *Listeria* phages are capable of generalized transduction, that is they more or less randomly package host DNA and can therefore transduce functional genetic markers into other cells (25). Tranduction frequencies range from 10^{-4} up to 5×10^{-2}, depending on the phage and host used. The ability to package non-phage DNA appears to correlate with the genome structure of the viruses: the terminally redundant A118 does transduce, whereas the *cos*-site phage PSA does not (25). This correlation seems to be the case for other *Listeria* phages as well (unpublished data) and is likely dependent on the different DNA packaging mechanisms employed by these viruses, which were shown to have unrelated terminase enzymes.

Virulent *Listeria* Phages

As mentioned above, most *Listeria* phages are temperate and, with the exception of A511, nothing is known about the few virulent phages for the genus. Phage A511, however, is a particularly interesting virus. It has a very wide host range and can infect approximately 95% of the relevant *L. monocytogenes* strains found to be implicated in foodborne disease. Of its large genome, only a fraction has yet been analyzed in detail, namely the lysis gene region (44) and a 10 kb fragment containing most of the morphopoietic genes (40). A useful finding were the powerful promoter sequences controlling expression of the major late genes, *cps* and *tsh*, which enabled design of a reporter phage vehicle (see below).

Relationships and Phage Evolution

Comparative genomics demonstrates that phage A118 is only very distantly related to phage PSA (figure 37-2), but highly similar to the cryptic phage, ϕEGDe, found in the chromosome of *L. monocytogenes* EGDe (20). Most differences between A118 and ϕEGDe are found in the early gene region, whereas the late gene region is, with the exception of the major capsid genes, extremely conserved. Only a single gene, encoding a part of the virus tail structure, is conserved among these three *Listeria*-phage genomes. In addition, these three phages were shown to contain portions resembling functional regions of other phage genomes, in particular those infecting lactic acid bacteria and other members of the low G-C content subbranch of Gram-positive eubacteria (36, 70).

Of interest is the relatedness of the *Listeria* phage A511 to *Staphylococcus aureus* phage Twort. Both belong to a group of morphologically basically indistinguishable, obligately virulent myoviruses that infect various Gram-positive hosts (21). Phage A511 was reported to have significant homologies in its late gene region to an intron-containing sequence of Twort (28), raising the possibility of the presence of self-splicing introns in *Listeria* phages. Also, in contrast to the situation of the temperate phages, it seems more difficult to explain the presence of almost identical sequence elements in viruses that do not exist in a prophage state, and do not infect a common host. We have previously observed an unusually high rate of recombination in A511 (39), which suggests that these (and other) viruses may use some specially adapted mechanisms that augment their ability to participate in the genetic mix-and-match game.

Additional indications exist that point to a relatedness of *Listeria* phages to viruses of other closely related Gram-positive bacteria, such as *Brochothrix thermosphacta*. Several short regions of high DNA homology were identified in morphologically unrelated phages of the two genera (unpublished data). One such region contains an *ssb* gene that is identical in the viruses of different origin but is flanked by unrelated portions of the genomes, which may reflect a good example of a limited, modular exchange.

Taken together, even the limited data on *Listeria* phages which are available to date clearly support the “mosaics” model of phage genome building, where phage genomes are built from genetic modules. Functional segments are accessible through different mechanisms from a large gene pool (23, 24). Horizontal exchange in phages is obviously dependent on the genetic material of their hosts, which may restrict promiscuous exchange. Nevertheless, *Listeria* phage genomic analyses add to the growing evidence that individual phages likely have evolved from a limited number of ancestral phages. This suggests a divergent evolution of bacterial viruses which is strongly influenced by continuous adaptation and genetic exchange of significant portions of their genomes (see chapters 4 and 27 for further discussion of these concepts).

Applications

Typing Phage

Various systems have been reported for phage typing of *Listeria* (4, 6, 15, 16, 19, 31, 32, 45, 52, 53, 64). Typability (sensitivity to at least one phage of a given set) is heavily dependent on the bacteria: serovar 3 strains are mostly resistant, in contrast to the high phage sensitivity of serovar 4 strains. Virulent phage with a broad lytic range such as A511 increased overall typability from around 70% to more than 90% (32, 65). Currently this technique provides the simplest and most widely used *Listeria* typing method (45), and it provides a sensitive means for tracing the origin and course of foodborne outbreaks of listeriosis.

Reporter Phage

The potential of genetically engineered phage is widely acknowledged in cloning procedures involving well-characterized bacteriophages such as phage λ. Because of its broad host range, phage A511 was selected as a candidate for construction of a reporter bacteriophage. A genetic fusion of *Vibrio harveyi luxA* and *luxB* genes was introduced into the A511 genome, under control of the powerful *cps* promoter (39). Following infection of *Listeria* cells by A511::*luxAB*, viral gene expression results in bioluminescent bacteria which can easily be detected and quantitatively monitored even in a mixed bacterial flora. A thorough evaluation of the system confirmed its usefulness as a quick and sensitive method for detection of viable *Listeria* in a variety of foods (40). See chapter 46 for further discussion of this reporter phage approach to bacteria identification and diagnosis.

Killer Phage

The potential usefulness of phage in bio-disinfection measures against *Listeria monocytogenes* on solid surfaces and production equipment was reported (56). However, this application faces many potential problems, and much more research is needed with respect to the potential of phage for eradicating *Listeria* in immensely complex environments such as food and feed. See chapter 48 of additional consideration of the use of intact phages as antibacterial agents.

Lytic Enzymes

Listeria phage endolysins can be produced in *E. coli*, and the recombinant enzymes retain high activity after affinity purification (42, 43). The enzymes evolved to exhibit stringent substrate specificity, that is they only lyse the *Listeria* cell peptidoglycan, with very few exceptions among closely related Gram-positive bacteria. Although the virus uses them from within the cell, Ply enzymes work equally well when added exogenously. A tiny amount of enzyme is sufficient to clear a dense suspension of *Listeria* cells within seconds. The enzymes are active in a pH range from 6 to 10, they are insensitive to common protease inhibitors and chelating agents, and moderate concentrations of detergents even increase their activity. Based on these properties, they have found a number of applications, such as rapid in vitro lysis (13, 41, 42, 44), removal of extracellular bacteria in eukaryotic cell invasion assays (58), selective release of intracellular metabolites such as ATP (59), and programmed self-destruction of intracellular *Listeria* cells with in the cytosol of macrophages (14). A novel approach for biological control of *Listeria monocytogenes* in fermented milk products is the production and secretion of N-terminally modified Ply511 by recombinant, lactose-utilizing *Lactococcus lactis* (18).

Another very interesting possibility is the use of the recombinantly produced cell wall binding domains of these enzymes (see above). The high-affinity, specific CBD polypeptides can be used for immobilization of host cells on solid surfaces such as coated microplates or magnetic beads. When fused to a fluorescent label, they have properties similar to a labeled antibody and allow specific decoration and efficient separation of *Listeria* cells from mixed bacterial populations and even within infected bacterial cells (30, 37).

Integration Vectors

On the basis of cloned integrase genes, two chromosomal integration vectors have been constructed. The first uses the integrase and phage attachment site of the A118-related phage, U153, and the second uses the analogous elements of phage PSA (29). These vectors are propagated in *E. coli* and can be transferred to *L. monocytogenes* by conjugation or by electroporation. Since these plasmids cannot replicate in *L. monocytogenes*, retention of the drug-resistant phenotype requires chromosomal integration. Using the U153*int*-based vector to integrate the genes for Listeriolysin O (LLO) and ActA, Lauer and coworkers showed that these genes are expressed well from the *comK* attachment site (29). Using the PSA vector, it has been shown that the *secA2* gene can be expressed well from the tRNA$^{\mathrm{Arg}}$ attachment site (30).

Transducing Phage

A recent and important finding was the use of *Listeria* phage for generalized transduction of genetic material from one strain to others (25). Of particular use is the introduction of marker-tagged mutations and associated phenotypes into a clean genetic background, which enables detailed genetic mapping and characterization of the mutation. The most useful phages for this purpose are P35 and U153 (serovar 1/2 strains), and A500 (serovar 4 strains). Phages infecting *L. innocua* and *L. ivanovii* were not tested in this study, but it is conceivable that many of the temperate viruses infecting these species will also be generalized transducers.

Acknowledgments

Many thanks are due to past and present members of our laboratories, who contributed with their work and enthusiasm to the many aspects of the study of *Listeria* phages. We thank Carlos Canchaya for help in preparing figure 37-2, and Julia Dorscht and Kris Srinivasan for providing unpublished information. Past research at TU Munich (M. J. L.) was supported by the Bayerisches Staatsministerium für Ernährung, Landwirschaft und Forsten (Germany), the European Union, the HFSPO (France), the Alexander von Humboldt Foundation (Germany), the Freunde und Förderer des FML Weihenstephan, the BMBF PathoGenoMik Competence Network, and by partners from the industry.

References

1. Ackermann, H.-W. 1999. Tailed bacteriophages: the order caudovirales. Adv. Virus Res. 51:135–201.
2. Ackermann, H.-W., A. Audurier, and J. Rocourt. 1981. Morphologie de bacteriophages de *Listeria* monocytogenes. Ann. Virol. 132:371–382.
3. Ackermann, H.-W., and M. S. DuBow. 1987. *Listeria* phages, pp. 108–113. *In* Viruses of Procaryotes, vol. 2, Natural Groups of Bacteriophages. CRC Press, Boca Raton, Fla.

4. Audurier, A., R. Chatelain, F. Chalons, and M. Pichaud. 1979. Lysotypie de 823 souches de *Listeria monocytogenes* isolées en France de 1958 à 1978. Ann. Microbiol. 130:179–189.
5. Audurier, A., J. Rocourt, and A. L. Courtieu. 1977. Isolement et caractérisation de bacteriophages de *Listeria monocytogenes*. Ann. Microbiol. 128:185–188.
6. Audurier, A., A. G. Taylor, B. Carbonelle, and J. McLauchlin. 1984. A phage typing system for *Listeria monocytogenes* and its use in epidemiological studies. Clin. Invest. Med. 7:229–232.
7. Bradley, D. E. 1967. Ultrastructure of bacteriophage and bacteriocins. Bacteriol. Rev. 31:230–314.
8. Bujnicki, J. M., and M. Radlinska. 2001. Cloning and characterization of M.*Lmo*A118I, a novel DNA:m^{4}C methyltransferase from the *Listeria monocytogenes* phage A118, a close homolog of M.*Ngo*MXV. Acta Microbiol. Pol. 50:155–160.
9. Chiron, J. P., P. Maupas, and J. L. Bind. 1975. Phagic induction in *Listeria monocytogenes*, pp. 271–279. *In* M. Woodbine (ed.) Problems of Listeriosis. Leicester University Press, Leicester, UK.
10. Chiron, J. P., P. Maupas, and F. Denis. 1977. Ultrastructure des bacteriophages de *Listeria monocytogenes*. C. R. Soc. Biol. (Paris) 1171:488–491.
11. Clark, E. E., L. Wesley, F. Fiedler, N. Promadej, and S. Kathariou. 2000. Absence of serotype-specific surface antigen and altered teichoic acid glycosylation among epidemic-associated strains of *Listeria monocytogenes*. J. Clin. Microbiol. 38:3856–3859.
12. Curtis, G. D. W., and R. G. Mitchell. 1992. Bacteriocin (monocin) interactions among *Listeria monocytogenes* strains. Int. J. Food. Microbiol. 16:283–292.
13. Dhar, G., K. F. Faull, and O. Schneewind. 2000. Anchor structure of cell wall surface proteins in *Listeria monocytogenes*. Biochemistry 39:3725–3733.
14. Dietrich, G., A. Bubert, L. Gentschev, Z. Sokolovic, A. Simm, A. Catic, S. H. E. Kaufmann, J. Hess, A. A. Szalay, and W. Goebel. 1998. Delivery of antigen-encoding plasmid DNA into the cytosol of macrophages by attenuated suicide *Listeria monocytogenes*. Nat. Biotechnol. 16:181–185.
15. Estela, L. A., and J. N. Sofos. 1993. Comparison of conventional and reversed phage typing procedures for identification of *Listeria* spp. Appl. Environ. Microbiol. 59:617–619.
16. Estela, L. A., J. N. Sofos, and B. B. Flores. 1992. Bacteriophage typing of *Listeria monocytogenes* cultures isolated from seafoods. J. Food Protect. 55:13–17.
17. Fiedler, F., and G. J. Ruhland. 1987. Structure of *Listeria monocytogenes* cell walls. Bull. Inst. Pasteur 85:287–300.
18. Gaeng, S., S. Scherer, H. Neve, and M. J. Loessner. 2000. Gene cloning and expression and secretion of *Listeria monocytogenes* bacteriophage-lytic enzymes in *Lactococcus lactis*. Appl. Environ. Microbiol. 66:2951–2958.
19. Gerner-Smidt, P., V. T. Rosdahl, and W. Frederiksen. 1993. A new Danish *Listeria monocytogenes* phage typing system. APMIS 101:160–167.
20. Glaser P., L. Frangeul, C. Buchrieser, C. Rusniok, A. Amend, F. Baquero, P. Berche, H. Bloecker, P. Brandt, T. Chakraborty, A. Charbit, F. Chetouani, E. Couvé, A. de Daruvar, P. Dehoux, E. Domann, G. Domínguez-Bernal, E. Duchaud, L. O. Durant, Dussurget, K.-D. Entian, K. H. Fsihi, F. Garcia-Del Portillo, P. Garrido, L. Gautier, W. Goebel, N. Gómez-López, T. Hain, J. Hauf, D. Jackson, L.-M. Jones, U. Kaerst, J. Kreft, M. Kuhn, F. Kunst, G. Kurapkat, E. Madueño, A. Maitournam, J. Mata Vicente, E. Ng, H. Nedjari, G. Nordsiek, S. Novella, B. de Pablos, J.-C. Pérez-Diaz, R. Purcell, B. Remmel, M. Rose, T. Schlueter, N. Simoes, A. Tierrez, J.-A. Vázquez-Boland, H. Voss, J. Wehland, and P. Cossart. 2001. Comparative genomics of *Listeria* species. Science 294:849–852.
21. Jarvis, A. W., L. J. Collins, L., and H. W. Ackermann. 1993. A study of five bacteriophages of the Myoviridae family which replicate on different Gram-positive bacteria. Arch. Virol. 133:75–84.
22. Hamon, Y., and Y. Peron. 1966. Sur la nature des bacteriocines produites par *Listeria*. C. R. Acad. Sci. Ser. D 263:198–200.
23. Hendrix, R. W., M. C. M. Smith, R. N. Burns, M. E. Ford, and G. F. Hatfull. 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. Proc. Natl. Acad. Sci. USA 96:2192–2197.
24. Highton, P. J., Chang, Y., and R. J. Myers. 1990. Evidence for the exchange of segments between genomes during the evolution of lambdoid bacteriophages. Mol. Microbiol. 4:1329–1340.
25. Hodgson, D. 2000. Generalized transduction of serotype 1/2 and serotype 4b strains of *Listeria monocytogenes*. Mol. Microbiol. 35:312–323.
26. Jasinska, S. 1964. Bacteriophages of lysogenic strains of *Listeria monocytogenes*. Acta Microbiol. Pol. 13:29–44.
27. Kalmokoff, M. L., and J. Farber. 1999. Bacteriocin-like activities among various species of *Listeria*. Int. J. Food Microbiol. 50:191–201.
28. Landthaler, M., and D. A. Shub. 1999. Unexpected abundance of self-splicing introns in the genome of bacteriophage Twort: introns in multiple genes, a single gene with three introns, and exon skipping by group I ribozymes. Proc. Natl. Acad. Sci. USA 96:7005–7010.
29. Lauer, P., M. Y. N. Chow, M. J. Loessner, D. A. Portnoy, and R. Calendar. 2002. Construction, characterization, and use of two *Listeria monocytogenes* site-specific phage integration vectors. J. Bacteriol. 184:4177–4186.
30. Lenz, L. L., and D. A. Portnoy. 2002. Identification of a second *Listeria secA* gene associated with protein secretion and the rough phenotype. Mol. Microbiol. 45:1043–1056.
31. Loessner, M. J. 1991. Improved procedure for bacteriophage typing of *Listeria* strains and evaluation of new phages. Appl. Environ. Microbiol. 57:882–884.
32. Loessner, M. J., and M. Busse. 1990. Bacteriophage typing of *Listeria* species. Appl. Environ. Microbiol. 56:1912–1918.
33. Loessner, M. J., L. Estela, R. Zink, and S. Scherer. 1994. Taxonomical classification of 20 newly isolated *Listeria* bacteriophages by electron microscopy and protein analysis. Intervirology 37:31–35.
34. Loessner, M. J., S. Goeppl, and M. Busse. 1991. Comparative inducibility of bacteriophage in naturally lysogenic and lysogenized strains of *Listeria* spp. by UV light and mitomycin C. Lett. Appl. Microbiol. 12:196–199.

35. Loessner, M. J., S. Goeppl, and M. Busse. 1991. The phagovar variability of *Listeria* strains under the influence of virulent and temperate bacteriophages. Lett. Appl. Microbiol. 12:192–195.
36. Loessner, M. J., R. B. Inman, P. Lauer, and R. Calendar. 2000. Complete nucleotide sequence, molecular analysis and genome structure of bacteriophage A118 of *Listeria monocytogenes*: implications for phage evolution. Mol. Microbiol. 35:324–340.
37. Loessner, M. J., K. Kramer, F. Ebel, and S. Scherer. 2002. C-terminal domains of *Listeria* bacteriophage murein hydrolases determine specific recognition and high affinity binding to the bacterial cell wall carbohydrates. Mol. Microbiol. 44:335–349.
38. Loessner, M. J., I. B. Krause, T. Henle, and S. Scherer. 1994. Structural proteins and DNA characteristics of 14 *Listeria* typing bacteriophages. J. Gen. Virol. 75:701–710.
39. Loessner, M. J., C. E. D. Rees, G. S. A. B. Stewart, and S. Scherer. 1996. Construction of luciferase reporter bacteriophage A511::*luxAB* for rapid and sensitive detection of viable *Listeria* cells. Appl. Environ. Microbiol. 62:1133–1140.
40. Loessner, M. J., M. Rudolf, and S. Scherer. 1997. Evaluation of luciferase reporter bacteriophage A511::*luxAB* for detection of *Listeria monocytogenes* in contaminated foods. Appl. Environ. Microbiol. 63:2961–2965.
41. Loessner, M. J., and S. Scherer. 1995. Organization and transcriptional analysis of the *Listeria* phage A511 late gene region comprising the major capsid and tail sheath genes *cps* and *tsh*. J. Bacteriol. 177:6601–6609.
42. Loessner, M. J., A. Schneider, and S. Scherer. 1995. A new procedure for efficient recovery of DNA, RNA, and proteins from *Listeria* cells by rapid lysis with a recombinant bacteriophage endolysin. Appl. Environ. Microbiol. 61:1150–1152.
43. Loessner, M. J., A. Schneider, and S. Scherer. 1996. Modified *Listeria* bacteriophage lysin genes *ply* allow efficient overexpression and one-step purification of biochemically active fusion proteins. Appl. Environ. Microbiol. 62:3057–3060.
44. Loessner, M. J., G. Wendlinger, and S. Scherer. 1995. Heterogeneous endolysins in *Listeria monocytogenes* bacteriophages: a new class of enzymes and evidence for conserved holin genes within the siphoviral lysis cassettes. Mol. Microbiol. 16:1231–1241.
45. McLauchlin, J., A. Audurier, A. Frommelt, P. Gerner-Smidt, Ch. Jacquet, M. J. Loessner, N. van der Mee-Marquet, J. Rocourt, S. Shah, and D. Wilhelms. 1996. WHO study on subtyping of *Listeria monocytogenes*: results of phage typing. Int. J. Food Microbiol. 32:289–299.
46. Mereghetti, L., R. Quentin, N. Marquet-Van der Mee, and A. Audurier, A. 2000. Low sensitivity of *Listeria monocytogenes* to quaternary ammonium compounds. Appl. Environ. Microbiol. 66:5083–5086.
47. Ortel, S. 1978. Untersuchungen ueber Monocine. Zentralbl. Bakteriol. Hyg. Abt. I, Ser. A, 242:72–78.
48. Ortel, S. 1981. Lysotypie von *Listeria monocytogenes*. Z. Ges. Hyg. 27:837–840.
49. Ortel, S., and H.-W. Ackermann. 1985. Morphologie von neuen Listeriaphagen. Zentralbl. Bakteriol. Hyg. Abt. I, Ser. A, 260:423–427.
50. Rocourt, J. 1986. Bacteriophages et bacteriocines du genre *Listeria*. Zentralbl. Bakteriol. Hyg. Abt. I, Ser. A, 261:12–28.
51. Rocourt, J., H.-W. Ackermann, M. Martin, A. Schretten brunner, and H. P. R. Seeliger. 1983. Morphology of *Listeria innocua* bacteriophages. Ann. Virol. 134E:245–250.
52. Rocourt, J., A. Audurier, A. L. Courtieu, J. Durst, S. Ortel, A. Schrettenbrunner, and A. G. Taylor. 1985. A multi-centre study on the phagetyping of *Listeria monocytogenes*. Zentralbl. Bakteriol. Hyg. Abt. I, Ser. A, 259:489–497.
53. Rocourt, J., B. Catimel, and A. Schrettenbrunner. 1985. Isolement de bacteriophages de *Listeria seeligeri* et *L. welshimeri*. Lysotypie de *L. monocytogenes, L. ivanovii, L. innocua, L. seeligeri, L. welshimeri*. Zentralbl. Bakteriol. Hyg. Abt. I, Ser. A 259:341–350.
54. Rocourt, J., M. Gilmore, W. Goebel, and H. P. R. Seeliger. 1986. DNA relatedness among *Listeria monocytogenes* and *L. innocua* bacteriophages. Syst. Appl. Microbiol. 8:42–47.
55. Rocourt, J., A. Schrettenbrunner, and H. P. R. Seeliger. 1982. Isolation of bacteriophages from *Listeria monocytogenes* serovar 5 and *L. innocua*. Zentralbl. Bakteriol. Hyg. Abt. I, Ser. A 251:505–511.
56. Roy, B., H. W. Ackermann, S. Pandian, G. Picard, and J. Goulet. 1993. Biological inactivation of adhering *Listeria monocytogenes* by *Listeria* phages and a quaternary ammonium compound. Appl. Environ. Microbiol. 59:2914–2917.
57. Schultz, E. W. 1945. *Listerella* infections: a review. Stanford Med. Bull. 3:135–151.
58. Smith, M. C. M., and C. E. D. Rees. 1999. Exploitation of bacteriophages and their components, pp. 97–132. *In* Smith, M. C. M. and Sockett, R. E. (ed.) Methods in Microbiology, vol. 29. Academic Press, San Diego.
59. Stewart, G. S. A. B., M. J. Loessner, and S. Scherer. 1996. The bacterial *lux* gene bioluminescent biosensor revisited. ASM News 62:297–301.
60. Subramanya, H. S., V. Arciszewska, R. A. Baker, L. E. Bird, D. J. Sherratt, and D. B. Wigley. 1997. Crystal structure of the site-specific recombinase, XerD. EMBO J. 16:5178.
61. Sword, C. P., and M. J. Pickett. 1961. The isolation and characterization of bacteriophages from *Listeria monocytogenes*. J. Gen. Microbiol. 25:241–248.
62. Tran, H. L., F. Fiedler, D. A. Hodgson, and S. Kathariou. 1999. Transposon-induced mutations in two loci of *Listeria monocytogenes* serotype 1/2a result in phage resistance and lack of *N*-acetylglucosamine in the teichoic acid of the cell wall. Appl. Environ. Microbiol. 65:4793–4798.
63. Tubylewicz, H. 1963. Studies on the lysogeny of *Listeria monocytogenes* strains. Bull. Acad. Pol. Sci. Ser. Sci. Biol. 11:515–518.
64. Van der Mee-Marquet, N., M. J., Loessner, and A. Audurier. 1997. Evaluation of seven experimental phages for inclusion in the international phage set for epidemiological typing of *Listeria monocytogenes*. Appl. Environ. Microbiol. 63:3374–3377.

65. Vazquez-Boland, J. A., M. Kuhn, P. Berche, T. Chakraborty, G. Dominguez-Bernal, W. Goebel, B. Gonzalez-Zorn, J. Wehland, and J. Kreft. 2001. *Listeria* pathogenesis and molecular virulence determinants. Clin. Microbiol. Rev. 14:584–640.
66. Vukov, N., I. Moll, U. Blaesi, S. Scherer, and M. J. Loessner. 2003. Functional regulation of the *Listeria monocytogenes* bacteriophage A118 holin by an intragenic inhibitor lacking the first transmembrane domain. Mol. Microbiol. 48:173–186.
67. Vukov, N., S. Scherer, E. Hibbert, and M. J. Loessner. 2000. Functional analysis of heterologous holin proteins in a $\lambda\Delta S$ genetic background. FEMS Microbiol. Lett. 184:179–186.
68. Wendlinger, G., M. J. Loessner, and S. Scherer. 1996. Bacteriophage receptors on *Listeria monocytogenes* cells are the *N*-acetylglucosamine and rhamnose substituents of teichoic acids or the peptidoglycan itself. Microbiology 142:985–992.
69. Wilhelms, D., and D. Sandow. 1989. Preliminary studies on monocin typing of *Listeria monocytogenes* strains. Acta Microbiol. Hungarica 36:235–238.
70. Zimmer, M., E. Sattelberger, R. Inman, R. Calendar, and M. J. Loessner. 2003. Genome and proteome of *Listeria monocytogenes* phage PSA: an unusual case for programmed +1 translational frameshifting in structural protein synthesis. Mol. Microbiol. 50:303–317.
71. Zink, R., and M. J. Loessner. 1992. Classification of virulent and temperate bacteriophages of *Listeria* spp. on the basis of morphology and protein analysis. Appl. Environ. Microbiol. 58:296–302.
72. Zink, R., M. J. Loessner, I. Glas, and S. Scherer. 1994. Supplementary *Listeria*-typing with defective *Listeria* phage particles (monocins). Lett. Appl. Microbiol. 19:99–101.
73. Zink, R., M. J. Loessner, and S. Scherer. 1995. Characterization of cryptic prophages (monocins) in *Listeria* and sequence analysis of a holin/endolysin gene. Microbiology 141:2577–2589.

38

Mycobacteriophages

GRAHAM F. HATFULL

Mycobacteriophages are viruses of the Mycobacteria. The interest in these phages derives in large part from the medical significance and biological idiosyncrasies of their hosts. Mycobacteria are acid-fast staining bacteria with characteristic waxy cell walls that can be readily divided into two groups based on their growth rate: slow-growers such as *Mycobacterium tuberculosis* that have a doubling time of 24 hours and fast-growers such as *Mycobacterium smegmatis* with 3–4 hour doubling times (for reviews see 8, 32) Several mycobacterial species are important human and animal pathogens, with the most notorious being *M. tuberculosis* and *Mycobacterium leprae*, the causative agents of tuberculosis and leprosy, respectively (8). The extent of these diseases is alarming—*M. tuberculosis* is the leading cause of human mortality from a single infectious disease and the increased prevalence of multiple drug-resistant *M. tuberculosis* strains greatly complicates its treatment. A study of the mycobacteriophages offers potential for the development of novel methodologies for the diagnosis, prevention, and treatment of these diseases as well as revealing interesting biological features of their unusual bacterial hosts (30, 32, 33).

Our knowledge of the composition and diversity of the huge global population of bacteriophages, estimated at 10^{31} particles (11, 83), is meager and we are just beginning to learn how the mycobacteriophages are related to other phages in the population. While there is no reason to suppose that they are atypical, mycobacteriophages presumably have strategies for infection through complex mycobacterial cell walls, they must be adapted for growth in slow-growing hosts, and they could conceivably be involved in mycobacterial pathogenicity. Elucidating these features as well as numerous additional aspects of viral–host intimacy—DNA integration, gene expression, adsorption, lysis, etc.—suggests that mycobacteriophages offer a fruitful field for exploration.

Mycobacteriophages were first investigated in the 1950s, prompted by their utility in phage-typing of clinical mycobacterial isolates (68, 76). Over 200 different mycobacteriophages infecting a broad variety of mycobacterial hosts have been described, some with quite narrow host ranges (such as DS6A that infects only members of the *M. tuberculosis* complex) and others that infect a wide variety of slow- and fast-growing strains (33). Until the late 1980s, rather few had been studied in molecular detail, although significant progress has been made since and there are several helpful reviews (29–31, 33, 51). In this chapter, I will focus on recent studies, most notably in the burgeoning area of bacteriophage genomics.

Mycobacteriophage Genomics

The development of high-throughput DNA sequencing technologies has made it possible to determine readily the complete genome sequences of bacteria and bacteriophages. The first mycobacteriophage genome to be sequenced was that of L5 in 1993 (34) followed by D29 and TM4 in 1998 (22, 23) and Bxb1 in 2000 (53); more recently, the genomes of an additional 10 mycobacteriophages (Barnyard, Bxz1, Bxz2, Che8, Che9c, Che9d, Corndog, Cjw1, Omega, and Rosebush) have been determined (57). Maps of the phage genomes are available at the web site www.thebacteriophages.org, All these phages infect *M. smegmatis* although many also infect other mycobacterial species including slow-growers such as BCG and *M. tuberculosis*; partial sequence information is available for mycobacteriophages DS6A and Ms6 (24, 25). Genome sequences of several mycobacterial hosts have also been determined including two strains of *M. tuberculosis* (H37Rv and CDC1551) and *M. leprae* (13, 14, 21); sequencing of the *M. smegmatis*, *Mycobacterium avium*, *Mycobacterium bovis*, *Mycobacterium marinum*, and BCG genomes is in progress with completion expected soon.

Overview and Classification of Sequenced Mycobacteriophages

Comparative analysis of a group of genomes is awkward without some general sense of their overall relatedness.

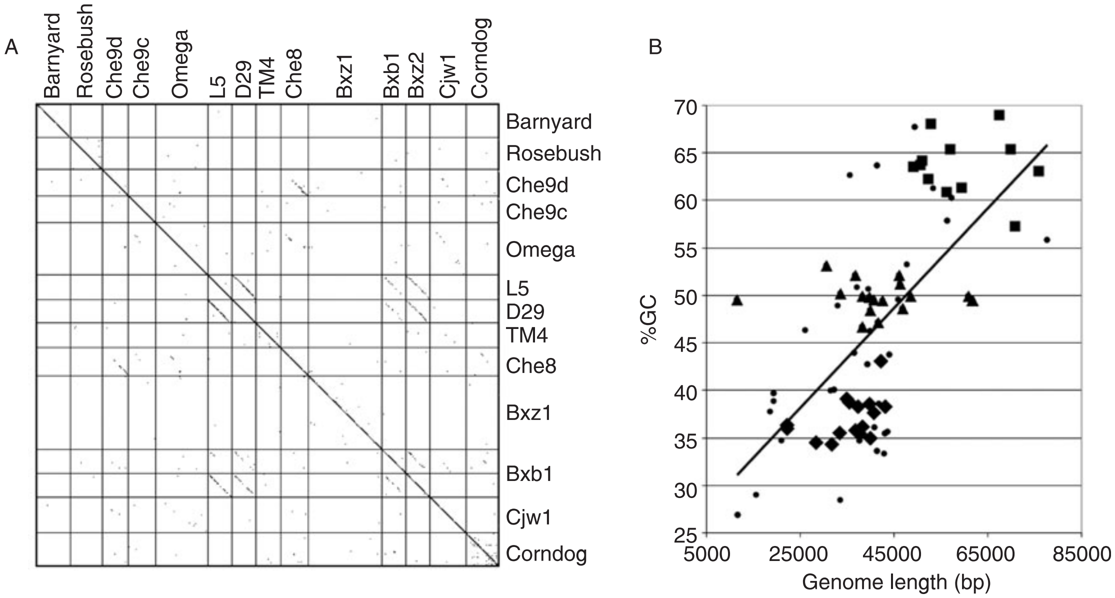


Figure 38-1 Mycobacteriophage genomic features. A: DNA sequence similarity among 14 completely sequenced mycobacteriophage genomes. B: Relationship between %GC content and genome length for all double stranded DNA phage genomes smaller than 100 000 base pairs. Phages are as follows: mycobacteriophages (■), *E. coli* and *Salmonella* phages (▲), dairy phages (♦), and others (●).

A hierarchical organization would be helpful but is problematic with double-stranded DNA phages in general and with the mycobacteriophages specifically. These difficulties are illustrated by the following general features of the mycobacteriophage genomes. First, at the nucleotide level they are quite diverse (figure 38-A). For example, some phages have significant levels of nucleotide similarity, with phages L5 and D29 being the most closely related pair; both phage Bxz2—and to a lesser extent phage Bxb1—also have nucleotide similarity to these phages. Likewise, phages Che9d and Che8 share some modest segments of nucleotide sequence similarity. In contrast, other phages—such as Barnyard, Rosebush, Che9c, Omega, Bxz1, Cjw1, and Corndog— have no extensive nucleotide sequence similarity to the other phages (figure 38-A).

Secondly, there is a substantial variation in genome length, although they are generally larger than other double-stranded DNA tailed phage genomes (table 38-1). The reason for this is not clear but a correlation between the GC% content and phage genome length has been noted (figure 38-1B) (57). The relatively large genome size can thus be considered in the context of their high GC%, rather than reflecting the necessity for additional functions needed to propagate within mycobacterial hosts.

Thirdly, there is no simple distinction between temperate and lytic viruses and examination of plaque morphologies does not provide a clear distinction between these classes. Some phages, such as L5, Bxb1, and Bxz2, form obviously turbid plaques from which stable lysogens can be isolated; these are indisputably temperate. Other phages, such as D29, form clear plaques from which lysogens cannot be recovered (22). While D29 is thus a lytic phage, this tells us little about its genome structure, and the sequence of D29 reveals that it is simply a recent derivative of an L5-like temperate parent (22). Most of the other phages form plaques that are lightly turbid or hazy in appearance, and we favor the interpretation that while perhaps competent to form lysogens (or pseudolysogens), either the establishment or maintenance of lysogeny is poor under laboratory conditions with the bacterial hosts employed. For example, preliminary observations suggest that phages Che8 and Cjw1 can both form *M. smegmatis* lysogens under certain conditions.

Fourthly, there is no obvious distinction between phages that use *cos*- and *pac*-type packaging systems as described for the dairy phages (10). Eleven of the 14 mycobacteriophages contain genomes with defined termini with short single-stranded DNA cohesive termini (table 38-1). However, we have not been able to identify defined genomic ends in Bxz1, Rosebush or Barnyard and we presume that they have circularly permuted genomes. Phage Bxz1 is curious in that it contains a region with a large stretch (>60) of G residues through which sequencing reactions fail, although this does not represent genome ends. Bxz1 has been shown to be a generalized transducing phage consistent with a *pac*-type packaging system; it is not known whether phages Rosebush and Barnyard are also generalized transducers

Fifth, there is considerable variation in the viral morphologies. Bxz1 is the only phage in this group with

Table 38-1 General Features of 14 Mycobacteriophages and their Genomes

Phage	Genome size (bp)	G+C%	tRNAs (no)	ORFS (no)	Av. ORFsize (bp)	Novel genes No. (%)	Genome ends (no. of bases, polarity)	Head size (nm)	Tail length (nm)
L5	52,297	62.3	3	87	601	14 (16.1)	9 base, 3′	~60	140
D29	49,136	63.5	5	77	638	10 (13.0)	9 base, 3′	~60	140
TM4	52,797	68.1	0	92	574	53 (57.6)	10 base, 3′	~60	200
Bxb1	50,550	63.7	0	86	588	29 (33.7)	9 base, 3′	~60	130
Bxz1	156,102	64.8	28	225	694	164 (72.9)	Circular. Permuted	~80	78
Che8	59,471	61.3	0	112	531	20 (17.9)	10 base, 3′	~60	186
Bxz2	50,913	64.2	3	86	599	22 (25.6)	10 base, 3′	~60	140
Cjw1	75,931	63.1	1	141	546	77 (54.6)	9 base, 3′	~60	250
Corndog	69,777	65.4	0	122	572	63 (51.6)	9 base 3′	23 × 138	265
Che9c	57,050	65.4	0	84	671	39 (45.3)	10 base, 3′	35 × 80	165
Omega	110,857	61.4	2	237	466	134 (56.5)	4 base, 3′	~60	205
Che9d	56,276	60.9	0	111	507	38 (34.2)	10 base, 3′	~60	130
Barnyard	70,797	57.3	0	109	650	93 (85.3)	Circular permuted	~60	265
Rosebush	67,480	69.0	0	90	750	65 (72.2)	Circular permuted	~60	245
Total	979,434		42	1559		821 (52.7)			
Average	69,960	63.6	3	118.6	599	58.6			

a contractile tail and all the others have long flexible tails of varying lengths (table 38-1). Twelve of the phages have isometric heads, all of which are approximately 60 nm in diameter except for Bxz1, which is about 80 nm in diameter. Two of the phages—Che9c and Corndog—have unusual prolate heads (see chapter 2 for a discussion of phage classification and morphology).

Lastly, no hierarchical relationships among these phages can be reconstructed using protein sequence similarities since the genomes are highly mosaic (see below for further discussion). Thus, although Bxz1 may superficially appear to be different from the other phages in that it is substantially larger (156 kbp) and apparently differs in its genome organization, it cannot be easily taxonomically separated from the others. For example, it contains more genes in common with phage Corndog than are shared between phages Che9c and L5 (57). Thus, neither morphology, genome length, packaging style, life-style nor sequence similarity offers a simple classification scheme for these phages.

Architectural Features of Mycobacteriophage Genomes

There is no single architectural design unifying all 14 mycobacteriophage genomes although common features are apparent. For example, all the genomes are replete with protein-coding genes with few noncoding spaces. Large groups of genes are usually transcribed together in the same direction and are presumably cotranscribed from a relatively small number of phage promoters. However, there is no common transcriptional organization shared by all the phages (figure 38-2). With the exception of phage Bxz1, a cluster of genes (encompassing approx. 20 kbp) involved in virion structure and assembly can be recognized which are located in the left part of the genomes; this cluster is usually further divided into the DNA packaging functions (terminase), head genes, and tail genes, (figure 38-2). In four of the phages (TM4, Che8, Che9c, and Che9d), the terminase gene is located within 1 kbp of the genome termini—the prototypical arrangement seen in phage λ — but in general these parts of the genomes are quite variable and additional genes are often present. For example, four of the genomes (L5, D29, Bxb1, Bxz2) have lysis genes in this region rather than downstream of the structural gene cluster (as in Corndog, Che8, Che9c, Che9d, TM4, Cjw1, Barnyard, Omega, and Rosebush); in phage Corndog the terminase gene and genome left end are separated by over 13 kbp and there are about 30 genes in this interval (figure 38-2).

Nine of the phages (L5, D29, Bxb1, Bxz2, Che8, Che9c, Che9d, Cjw1, Omega) encode an integrase, and in each case the gene is situated close to the center of the genome; where known, the *attP* site is adjacent to *int*. The distance between the left end of the integrase gene and the left end of the genome varies from 46.3% (phage L5) to 58.4% (phage Bxb1) even though there is more than a 2-fold difference in total genome length among the integrase-encoding genomes (table 38-1). The reason why the *attP*/integrase cassette should be centrally located—giving rise to similarly sized left and right arms—is not clear. We note, however, that the region between the *int* and the structural genes (or the lysis genes if they are in this location) is variable in size and composition (figure 38-2). In L5 they are almost adjacent and in Omega this interval is >9 kbp and contains approximately 30 genes, mostly of unknown function.

The organization and numbers of genes in the right parts of the genomes (to the right of the structural, lysis, and integration genes) varies greatly. In phage D29 this segment is approximately 22 kbp and includes about 60 genes; in contrast, phage Omega has more than 130 genes in a span of over 55 kbp, more than the entire D29 genome! Typically these regions contain a preponderance of small open reading frames, most of which do not match previously described genes although genes involved in DNA metabolism, replication, recombination, and regulation can be recognized (see below).

Phage-Encoded tRNA Genes

Six of the sequenced mycobacteriophages encode tRNA genes (table 38-1). In phages L5, D29, and Bxz2 there is a small cluster of tRNA genes located approximately 4 kbp from the left genome end, upstream of the terminase genes. Phages Omega and Cjw1 each contain one standard tRNA gene close to the right end of the genome, plus one nonstandard tRNA gene; the nonstandard Omega tRNA has eight bases in the anticodon loop, suggesting it may be a frameshift suppressor. Bxz1 contains the largest number of tRNA genes (twenty-eight) arranged in two large, loosely organized clusters, of which 27 are typical and correspond to 16 different amino acid specificities; this is the largest set of tRNA genes identified in any viral genome. The one atypical tRNA has a 5′-CUA anticodon and could function as a suppressor of 5′-UAG stop codons, raising the question of whether UAG is used as a stop codon. Thirty-five of the 231 open reading frames terminate with UAG and although many of these are followed by short noncoding segments, alternative stop codons, or adjacent downstream coding regions in the same frame, this is not true for all these genes. Thus, while the putative suppressor tRNA could regulate the expression of specific genes, it may not confer a wholesale modification of the Bxz1-phage genetic code.

The role of the mycobacteriophage-encoded tRNAs is not known, although Kunisawa noted that phage D29 codons corresponding to the D29 tRNAs have higher relative usage than they do in the host (43). He concludes that the host (*M. tuberculosis*) has a paucity of the cognate tRNA for these codons and that the phage compensates by

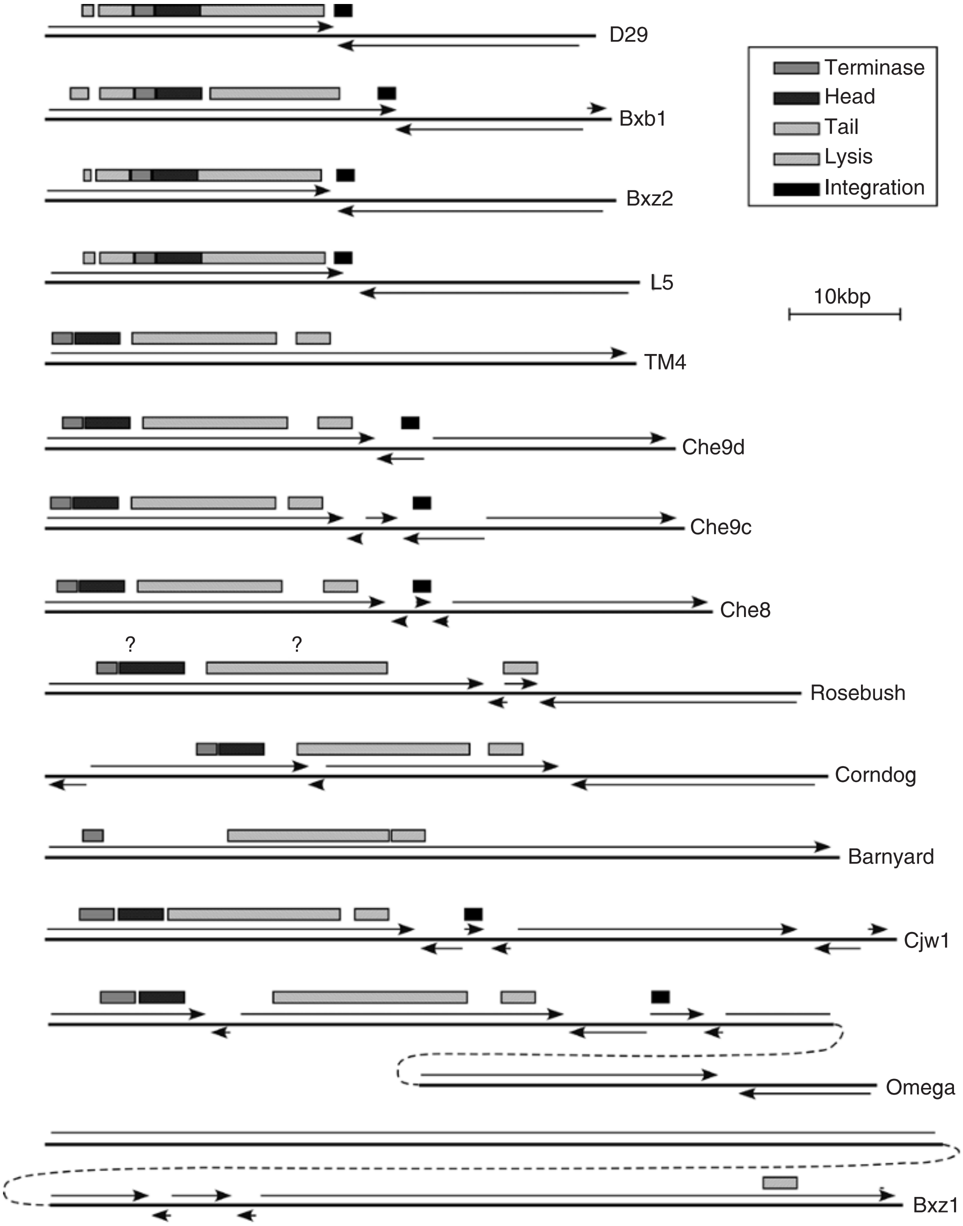


Figure 38-2 Mycobacteriophage genomic architectures. The genome of each of the 14 completely sequenced mycobacteriophage genomes is represented as a thick horizontal line with arrows indicating the direction of transcription. The locations of genes involved in integration, lysis, the head and tail genes, and packaging (terminase) are indicated.

carrying additional copies of the appropriate tRNA genes. However, no similar correlation is seen in Bxz1, where less than one half of the 27 Bxz1 tRNAs correspond to codons that are overabundant in the phage (relative to *M. tuberculosis*). Either this explanation is not generally applicable or the codon usage of the preferred host for Bxz1 is substantially different from that of *M. tuberculosis*. It is noteworthy that Bxz1 does not efficiently infect slow-growing mycobacterial strains and the codon frequencies in *M. smegmatis* have yet to be evaluated; phage D29 does, however, infect both fast- and slow-growing strains.

Virion Structure and Assembly

Capsid Proteins

N-terminal sequence analysis of phage L5 virion proteins shows that the capsid subunit is encoded by gene *17* and covalently crosslinked in the mature capsid, similarly to the coliphage HK97 capsid (34, 65); D29 gp17, Bxb1 gp14, Bxz2 gp17, Che9d gp7, and TM4 gp9 all share significant sequence similarity with L5 gp17 (table 38-2). While none of these phages are close relatives of the phage HK97,

Table 38-2 Mycobacteriophage Capsid Genes

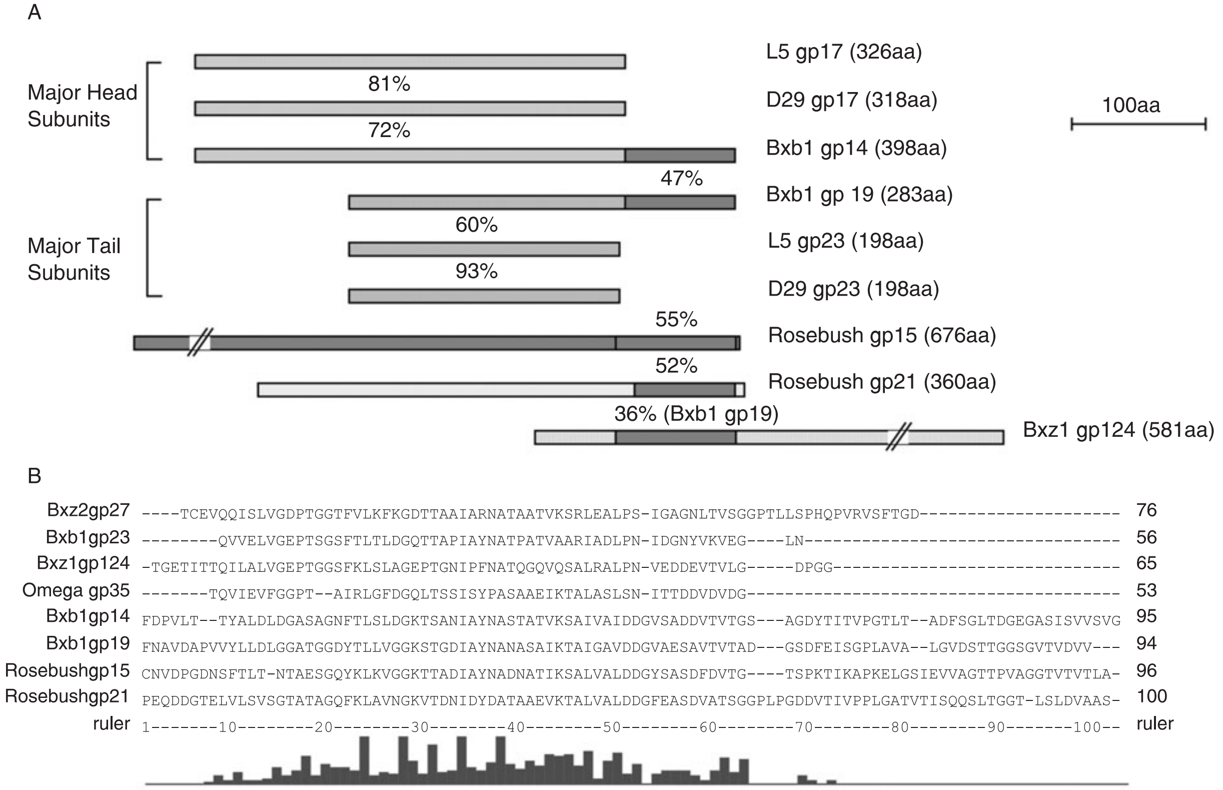
Phage	Capsid	Size (amino acids)	Capsid homologs	Other features
L5	gp17	327	D29 gp17, Bxb1 gp14, Bxz2 gp17, A118 gp6, Che9d gp7, TM4 gp9	Member of the HK97 gp5 family of capsid proteins
D29	gp17	319	L5 gp17, Bxb1 gp14, Bxz2 gp17, A118 gp6, Che9d gp7, TM4 gp9	Member of the HK97 gp5 family of capsid proteins
Bxb1	gp14	398	L5 gp17, D29 gp17, Bxz2 gp17, A118 gp6, Che9d gp7, TM4 gp9	Member of the HK97 gp5 family of capsid proteins. 180 amino acid C-terminal. extension at ends of Bxb1 gp19, Rosebush gp15, and gp21. In middle of Bxb1 gp23, Bxz1 g124, Omega gp35, and Bxz2 gp27
Bxz2	gp17	331	Bxb1 gp14, L5 gp17, D29 gp17, A118 gp6, TM4 gp9, Che9d gp7	Member of the HK97 gp5 family of capsid proteins
TM4	gp9	306	A118 gp6, Bxz2 gp17, D29 gp17, L5 gp17, Bxb1 gp14	Member of the HK97 gp5 family of capsid proteins
Che9d	gp7	312	R1t gp31, Bxz2 gp17, D29 gp17, L5 gp17, Bxb1 gp14	Member of the HK97 gp5 family of capsid proteins
Cjw1	gp12	498	BFK20 gp6, HK97 gp5, HK022 gp5, D3 gp6	Member of the HK97 gp5 family of capsid proteins
Omega	gp15	478	HK97 gp5, HK022 gp5, bIL285 gp44, (Bxb1 gp13?)	Member of the HK97 gp5 family of capsid proteins
Che9c	gp6	544	Rv1576c, Rv2650c, AgrC capsid, *Caulobacter* hyp., *M. loti* gp36, *H.* flu hyp.,SfV capsid, VWB gp2	Member of the HK97 gp5 family of capsid proteins
Che8	gp6?	274	DRA0099	PSI-BLAST connection to T3/T7 capsid subunits
Corndog	gp41?	290	HI0445/HI0446	HI0445/HI0446 is a homolog of ϕC31 gp36
Rosebush	gp15?	676		gp15 has C-terminal. extension of Bxb1 gp14 and gp19, and Rosebush gp21. No homologs.
Bxz1	??			Capsid gene identity uncertain; no homologs
Barnyard	??			Capsid gene identity uncertain; no homologs

PSI-BLAST analysis reveals them all to be members of the HK97 gp5 capsid-protein superfamily. Phage Cjw1 gp12 and phage Omega gp15 are more closely related to the phage HK97 gp5, and phage Cjw1 gp12 shares 23% identity with HK97 gp5. Interestingly, the prolate-headed phage Che9c encodes a protein (gp6) with strong sequence similarity to proteins Rv1576c and Rv2650c of the *M. tuberculosis* prophage-like elements ϕRv1 and ϕRv2, respectively, and all three can be drawn into the HK97 gp5 family by PSI-BLAST analysis. Particles corresponding to the ϕRv1 and ϕRv2 elements have yet to be identified but these observations raise the question as to whether these might also form prolate structures similar to phage Che9c heads.

The major capsid subunit gene is harder to identify in phages Che8, Corndog, Rosebush, Bxz1, and Barnyard. In phage Che8, a likely candidate is gene *6*, whose product is related (26% identity) to the DRA0099 protein that is encoded by a prophage element in *Deinococcus radiodurans*; PSI-BLAST analysis links DRA0099 and Che8 gp6 to capsid proteins of phages T3, T7, and cyanophage P60 (table 38-2). In Corndog, gp41 is a likely capsid subunit since it is related to a gene product in a genetic island in *Haemophilus influenzae* that in turn is related to phage ΦC31 gp36, a known member of the HK97 gp5 superfamily. The capsid subunits of phages Bxz1 and Barnyard have yet to be identified, although the Barnyard gene is most likely one of several (*18–28*) upstream of the tape measure gene. In phage Rosebush, gene *15* is a candidate capsid subunit gene for reasons that are described below.

The Unusual Capsid Subunit of Bxb1

Phage Bxb1 shares considerable overall similarity with phages L5 and D29, especially within the virion structure and assembly gene cluster. The major capsid subunits of Bxb1 (gp14), L5 (gp17) and D29 (gp17) are closely related; L5 gp17 and D29 gp17 share 82% identity, Bxb1 gp14 and L5 gp17 share 69% identity, and Bxb1 gp14 and D29 share 72% identity (figure 38-3). However, while L5 gp17 and D29 gp17 are similar lengths (326 and 318 amino acids, respectively), Bxb1 gp14 is larger (397 amino acids) due to a C-terminal extension of approximately 85 residues. Since the covalently crosslinked pentameric and hexameric forms of the Bxb1 capsid migrate more slowly during SDS-PAGE than those of L5 (and D29), it seems unlikely that this C-terminal appendage is cleaved off during assembly (53). Moreover, the role of the C-terminal extension is not obvious since clearly it is not required for L5 or D29 assembly. We note that the Bxz2 capsid subunit (gp17) is also closely related (56% identity with Bxb1 gp14) but lacks the C-terminal extension.



```
    Bxz2gp27 ----TCEVQQISLVGDPTGGTFVLKFKGDTTAAIARNATAATVKSRLEALPS-IGAGNLTVSGGPTLLSPHQPVRVSFTGD---------------------  76
    Bxb1gp23 --------QVVELVGEPTSGSFTLTLDGQTTAPIAYNATPATVAARIADLPN-IDGNYVKVEG----LN---------------------------------  56
   Bxz1gp124 -TGETITTQILALVGEPTGGSFKLSLAGEPTGNIPFNATQGQVQSALRALPN-VEDDEVTVLG-----DPGG------------------------------  65
  Omega gp35 -------TQVIEVFGGPT--AIRLGFDGQLTSSISYPASAAEIKTALASLSN-ITTDDVDVDG---------------------------------------  53
    Bxb1gp14 FDPVLT--TYALDLDGASAGNFTLSLDGKTSANIAYNASTATVKSAIVAIDDGVSADDVTVTGS---AGDYTITVPGTLT--ADFSGLTDGEGASISVVSVG  95
    Bxb1gp19 FNAVDAPVVYLLDLGGATGGDYTLLVGGKSTGDIAYNANASAIKTAIGAVDDGVAESAVTVTAD---GSDFEISGPLAVA--LGVDSTTGGSGVTVDVV---  94
Rosebushgp15 CNVDPGDNSFTLT-NTAESGQYKLKVGGKTTADIAYNADNATIKSALVALDDGYSASDFDVTG----TSPKTIKAPKELGSIEVVAGTTPVAGGTVTVTLA-  96
Rosebushgp21 PEQDDGTELVLSVSGTATAGQFKLAVNGKVTDNIDYDATAAEVKTALVALDDGFEASDVATSGGPLPGDDVTIVPPLGATVTISQQSLTGGT-LSLDVAAS- 100
       ruler 1-------10--------20--------30--------40--------50--------60--------70--------80--------90-------100-- ruler
```

Figure 38-3 The unusual structures of the Bxb1 capsid and major tail proteins. A: The capsid (major head subunit) of Bxb1 is similar to L5 and D29 capsids but contains a C-terminal extension that is related to a similar extension at the C-terminus of the Bxb1 major tail subunit. A similar element is also present at the C-termini of Rosebush gp15 and gp21; these may correspond to the major head and tail subunits of Rosebush, but the large N-terminal regions do not have database matches. Related segments are also found in a putative Bxz1 structural protein (gp124) and in the middle of both Bxb1 gp23 and Omega gp35 (not shown). The degree of amino acid sequence identity is shown for segments of vertically adjacent pairs. B: Amino acid sequence alignment of C-terminal extensions of Bxb1 proteins. Sequences were aligned using ClustalX. Residues present in all sequences are indicated by an asterisk above the alignment. Amino acid segments shown are as follows: Bxz2 gp27, 267–342; Bxb1 gp23, 393–448; Bxz1 gp124, 55–119; Omega gp35, 250–302; Bxb1 gp14, 303-397; Bxb1 gp19, 190-283; Rosebush gp15, 580-675; Rosebush gp21, 260–359. The function of this sequence motif is not known.

The extension at the C-terminus of Bxb1 gp14 is curious since a similar sequence is also present at the C-terminus of the Bxb1 major tail subunit, gp19 (figure 38-3). Bxb1 gp19 is a close relative of the major tail subunits of phages Bxz2 (gp23), D29 (gp23), and L5 (gp23), sharing 64%, 61%, and 60% identity respectively (figure 38-3). However, as with the capsid subunit, Bxb1 gp19 is longer than its counterparts in L5, Bxz2, and D29 due to an extension of about 85 residues at the C-terminus. Interestingly, this 85-residue segment is related to that at the end of Bxb1 gp14, with the two sharing 47% identity (figure 38-3). As with the capsid subunit, the Bxb1 major tail subunit is not processed (53). Thus, each Bxb1 particle is expected to contain about 600 copies of this motif, 415 in the capsid and about 200 in the tail!

This motif is quite promiscuous and is found in several other locations. For example, there is at least one more distantly related copy of this motif in the middle of putative minor tail proteins, Bxb1 gp23 and Bxz2 gp27; Bxz1 also contains a related sequence near the N-terminus of gp124, a putative structural protein. However, an especially striking relationship is in mycobacteriophage Rosebush, in which two open reading frames (*15* and *21*) each contain a copy of this motif at their C-termini (figure 38-3). The functions of these two proteins are not known but we speculate that they correspond to the capsid and major tail subunits, respectively.

Scaffold Proteins

Although the capsid proteins of coliphage HK97 and mycobacteriophage L5 are members of the same family of proteins, they differ in that HK97 gp5 contains a 105-residue N-terminal domain that plays a role in capsid assembly but which is proteolytically removed during capsid maturation. In contrast, L5 encodes a separate scaffold protein, gp16, which is a component of head-like particles in infected cells but absent from mature heads. Phages D29, Bxb1, and

Bxz2 — whose structural proteins are close relatives of L5 — also contain similar scaffold protein genes located immediately upstream of the capsid subunit gene. In phage TM4, the capsid subunit (gp9) is a similar size to L5 gp17 (305 versus 326 amino acids) and likely also has a separately encoded scaffold protein. While the upstream gene (gp8) is not obviously related to the L5/D29/Bxb1/Bxz2 scaffold proteins, PSI-BLAST analysis suggests that these are all members of a larger family of assembly proteins; mycobacteriophage Che8 gp5 and Che9d gp6 as well the lactococcal phage r1t gp30 also appear to be members of this group. In contrast, the putative capsid proteins of Che8 (gp6), Cjw1 (gp12), Corndog (gp41), Omega (gp15), and Che9c (gp6) are generally longer than L5 gp17 and do not have a scaffold-like gene immediately upstream; it seems probable, therefore, that these putative capsid proteins are structurally more similar to HK97 gp5 and utilize an N-terminal domain to promote capsid assembly.

Portals and Proteases

Genes encoding protease proteins involved in capsid assembly can also be identified in some of the mycobacteriophages, typically immediately upstream of the capsid assembly genes. For example, Che9c gp5, Omega gp14, and Corndog gp39 are weakly related to other phage proteases (e.g., ϕC31 gp35) as are proteins Rv1577c and Rv2651c encoded by the *M. tuberculosis* prophage-like elements ϕRv1 and ϕRv2. Other protease candidates are L5, gp15, D29 gp15, Bxz2 gp15, Bxb1 gp12, Che8 gp4, Cjw1 p11, Che9d gp5, and TM4 gp6, all of which are encoded between the capsid assembly and upstream putative portal genes. Putative portal proteins such as Omega gp13, Cjw1 gp10, Che9c gp4, and Corndog gp34 share sequence similarity with a large group of other phage portals (including HK97 gp3). L5 gp14, D29 gp14, Bxz2 gp14, Bxb1 gp11, Che9d gp4, TM4 gp5, Che8 gp3 form a second group of portal proteins that are related to ϕ31 gp5 and r1t gp27.

Tails

As noted above, all the sequenced mycobacteriophages—excepting Bxz1—have long, noncontractile tails. Tail length varies considerably from approximately 140 nm (L5) to about 260 nm (Rosebush) but in most cases the tails have a defined structure at the tip and a tail shaft composed of a number of rings. Side tail fibers have not been observed on any of these phages, although some (e.g., L5 and D29) have a visible spike protruding from the tail tip.

The visible shaft of these noncontractile tails is likely composed of rings of the major tail subunit. The genes encoding these proteins have been identified in several of the phages and they appear to fall into two main sequence groups: one group contains L5 gp23, D29 gp23, Bxb1 gp19, Bxz2 gp23, TM4 gp14, and Che9d gp14; the other contains Che9c gp12, Corndog gp49, Che8 gp11, Cjw1 gp18, and Omega gp31. The first group is related to virion subunits (presumably also major tail subunits) of some non-mycobacteriophage viruses such as lactococcal phage r1t (gp37) (79) and staphylococcal phage VWB (gp27) (1). One notable feature of these proteins is that they typically migrate somewhat slower than their predicted molecular weight in SDS polyacrylamide gel electrophoresis, an observation that has also been noted for the major tail subunit of phage λ.

N-terminal sequencing has identified several genes encoding minor tail subunits in phage L5 including the products of genes *6*, *26*, *27*, and *28* (34). Identifying gene *6* as a tail protein was somewhat surprising since it is encoded upstream of terminase and other structural genes; D29 gp6 and Bxz2 gp3 are homolog of L5 gp6 and encoded in a similar place in the genome. The product of L5 gene *26* is almost certainly the phage tape measure protein and is the largest gene in the L5 genome. L5 gp26 may be C-terminally processed since while the N-terminal sequence corresponds to the predicted start of the gene, the protein migrates significantly faster than its predicted molecular weight of 86 kDa. The two additional minor tail proteins, gp27 and gp28, may be tail tip components.

Programmed Translational Frameshifting in Tail Genes

It has been shown previously that phage λ encodes two gene products from the *G* and *T* genes (47). One is the expected gpG protein, the other (gpG-T, which is made at about 4% the amount of gpG) is expressed via a programmed −1 translational frameshift near the end of the *G* gene. Curiously, this frameshifting phenomenon within genes encoded just upstream of the tape measure protein is an extremely common feature among the double-stranded DNA tailed phages (84), although in some phages (e.g., Mu) it is a −2 frameshift (54). Genes expressed by similar frameshifting events can be identified in many of the mycobacteriophages, and in mycobacteriophage L5 the expression of gp24-25 via a −1 frameshift within the end of gene *24* has been experimentally demonstrated (84). It is predicted that phages D29, Bxb1, TM4, Bxz2, Bxz1, and Cjw1 also express tail genes via a −1 frameshift (table 38-3) and, with the exception of Bxz1, the encoded proteins are related, albeit weakly in some cases. Interestingly, in phages Che8, Corndog, Che9c, and Omega, genes in the same location are also predicted to be expressed via frameshifting, but in these cases a −2 frameshift is involved (table 38-3). Moreover, while phages Cjw1 *20* and *21* utilize a −1 frameshift, they are more closely related at the sequence level to the Che8/Corndog/Che9c/Omega group, suggesting that a switch between −1 and −2 frameshifting occurred relatively recently in their evolution.

Table 38-3 Frameshifts in Mycobacteriophage Tail Protein Genes

	"G"		"T"		
Phage	Gene	Homologs	Gene	Homologs	Shift
L5	24	D29 gp24; Bxz2 gp24; Bxb1 gp20	25	D29 gp25; Bxz2 gp25; Bxb1 gp21; TM4 gp16	−1
D29	24	L5 gp24; Bxb1 gp20; Bxz2 gp24	25	L5 gp25; Bxz2 gp25; Bxb1 gp21	−1
Bxb1	20	Bxz2 gp24; D29 gp24; L5 gp24	21	D29 gp25; L5 gp25; Bxz2 gp25	−1
Bxz2	24	Bxb1 gp20; D29 gp24; L5 gp24	25	Bxb1 gp21; D29 gp25; L5 gp25	−1
TM4	15	None	16	L5 gp25	−1
Bxz1	127	None	128	None	−1
Cjw1	20	Che8 gp12	21	Che8 gp13	−1
Che8	12	Corndog gp54; Omega gp32; Cjw1 gp20	13	Corndog gp55; Omega gp33	−2
Corndog	54	Che8 gp12; Omega gp32; Cjw1 gp20; Che9c gp13	55	Che8 gp13; Omega gp33; Cjw1 gp21	−2
Che9c	13	Omega gp32; Cjw1 gp20; Corndog gp54	14	None	−2
Omega	32	Corndog gp54; Cjw1 gp20; Che8 gp12; Che9c gp13	33	Corndog gp55; Che8 gp13	−2

"G" and "T" refer to the *G* and *T* genes of phage λ.

Tail Tape Measure Proteins

The genes encoding the mycobacteriophage tape measure proteins are easy to identify since they are typically the largest genes in the genome and the length of tape measure gene correlates with phage tail length (42). This relationship is also observed in the 13 mycobacteriophages with noncontractile tails although the tail length and tape measure genes vary considerably: ~140 nm (e.g., L5, D29, Bxb1) to over 260 nm (e.g., Corndog, Barnyard) and the corresponding genes from approximately 2.5 to 6 kbp. In a few cases (e.g., Barnyard, Omega, Che8) the gene is somewhat larger than would be predicted from the tail length and it is plausible that these proteins are proteolytically processed prior to their role in tail length determination (57).

The mycobacteriophage tape measure proteins have a generally similar composition that is high in alanine and glycine (combined amounts of 23% in L5 to 29.7% in Che8), and none have more than a single cysteine residue; the predicted pI is typically close to 8.0, ranging from 7.7 in Che9d to 8.3 in D29 and Rosebush. These proteins, however, are quite varied in their amino acid sequences and are related to each other in complex ways. The most closely related set of proteins are L5 gp26, D29 gp26, Bxb1 gp22, and Bxb2 gp26, which can be aligned over their entire lengths (pairwise identities range between 26% and 73%). However, a central portion of these proteins (~450 amino acids) also has similarity to a group of proteins encoded by streptococcal phages Sfi11 and O1205, and lactococcal phage r1t. In a more extreme case, Corndog gp57 has patches of similarity to a number of other mycobacteriophage tape measure proteins as follows: Omega gp34 over residues 11–472 (26%), 795–951 (28%), and 1337–1387 (40%); Che9c gp15 over residues 89–171 (33%), 763–954 (32%), and 1337–1411 (34%); Che8 gp14 over residues 719–954 (27%) and 1193–1390 (34%); Rosebush gp29 over residues 232–451 (26%); Barnyard gp33 over residues 223–505 (19%); Bxz2 gp26 over residues 237–428 (22%); Cjw1 gp22 over residues 733–904 (20%); and L5 gp26 and D29 gp26 over residues 295–428 (21%).

A particularly intriguing observation is the finding that the tape measure protein of Barnyard (gp33) includes a sequence motif related to bacterial cytokines (57). Previously it has been shown that *Micrococcus luteus* secretes a cytokine or resuscitation factor (Rpf) that promotes growth of dormant *M. luteus* cells (55); *Mycobacterium tuberculosis* contains five homologous genes of Rpf (Rv1009, Rv0867c, Rv2389c, Rv2450c, and Rv1884c) and there are three homologs in *M. leprae*, four in *Streptomyces coelicolor*, and two in *Corynebacterium glutamicum*. While these various proteins differ in size, they all share a well-conserved central segment of approximately 90 amino acids to which residues 1490–1584 of Barnyard gp33 are similar; the most closely related protein is Rv1009, which shares 61% identity over a 90-residue segment of Barnyard gp33. This sequence relationship strongly suggests a function of the tape measure protein that was not previously recognized—the ability to stimulate dormant host bacterial cells into a growing state to support successful phage infection.

Upon reflection, this activity should perhaps not have been unexpected. In their natural environments it is likely that most bacterial cells are dormant or in a nongrowing state. This poses a challenge to phages since these hosts are not expected to support a productive phage infection. The tape measure protein seems an ideal location to place this viral alarm clock since it is one of few phage proteins that must presumably pass through the membrane and into the cell prior to DNA injection through the tail. While little is known about how Rpf-mediated signaling works, the tape measure protein is well placed to mimic the resuscitation factor whether the signaling receptor is membrane-bound or cytoplasmic. Moreover, since the signaling motif is part of the phage particle, gene expression is not required to promote phage-mediated reawakening.

The ability to awake dormant bacterial hosts should confer a significant selective advantage to the virus and it might be expected to be a common viral strategy. While Barnyard is the only phage identified thus far that carries the Rpf motif, there are additional motifs within the tape measure proteins of other mycobacteriophages with similarity to other small, putatively secreted mycobacterial proteins. One of these motifs is found in the tape measure proteins of Che9c (gp15), Cjw1 (gp22), Rosebush (gp29), and Barnyard gp33, and is related to *M. tuberculosis* Rv1115. A third motif is present in phage TM4 (gp17), Che8 (gp14), and Omega (gp34) tape measures and is related to the putative secreted *M. tuberculosis* proteins, Rv0320 and Rv1728c. The functions of these *M. tuberculosis* proteins are not known, but we predict that they are also involved in cell–cell communication.

Imposters in Structural Gene Clusters

Lambdoid phages contain morons, open reading frames flanked by a promoter and a stem-loop terminator that are present in one genome but absent from related genomes; morons can be inserted in either orientation relative to the direction of transcription of the structural genes and may be expressed in lysogens (36, 40) (morons are further discussed in chapter 27). Identifying morons in mycobacteriophage genomes is tricky since mycobacterial promoters are less well defined and stem-loop terminators are rare. However, the structural gene cluster is a good place to search for both morons and gene insertions because of the well defined gene order.

The genomes of L5, D29, Bxb1, Bxz2, TM4, Cjw1, Che8, Che9d, and Che9c have a canonical organization of the structural genes with the only major departure being the location of the L5 gp6 tail gene as noted above. In both Corndog and Omega there are notable departures; in Omega there are approximately 15 open reading frames between the head and tail genes—two of which are transcribed on the opposite strand—where it is usual to find four to six genes involved in head–tail connection. Ten of the 15 have no database matches and three match Corndog genes present in a similar location (probably involved in head–tail connection). Interestingly, the two genes (*16* and *17*) immediately downstream of the capsid gene have similarities to glycosyl transferases and *O*-methyltransferases, respectively. This raises the intriguing possibility that the products of these genes are involved in post-translational modification of the phage particles.

In Corndog, a different departure in the structural gene organization is seen. Typically the portal and capsid genes flank the protease and scaffold, but in Corndog there are six genes in this region (genes *35–40*), one of which (gene *39*) encodes a putative protease. Two of the others (genes *37* and *38*) are homologs of Che8 gp109 and gp110, located at the right end of the Che8 genome, and Corndog gene *35* encodes an *O*-methyltransferase—a homolog of Che8 gp108—perhaps suggesting that Corndog (as well as Che8) particles may also post-translationally modified

Lysis Genes

Release of phage particles at the end of a lytic cycle is dependent on phage-encoded lysis genes (80). Identification and characterization of the mycobacteriophage lysis genes is of interest not only for understanding the mechanism and timing of lysis but also because of the possible therapeutic use of mycobacteriophage lysins; similar applications have shown promise for control of streptococcal and anthrax infections (20, 49, 74).

Putative lysis genes have been identified in mycobacteriophage Ms6, where three genes have been implicated: *lysA*, *ORF3*, and *hol* (25). Two of these, *lysA* and *ORF3*, have proposed enzymatic functions while *hol* encodes a putative holin (figure 38-4) (see chapter 10 for a discussion of the functional relationship between *lys* and *hol* genes). When the Ms6 LysA enzyme is expressed in *Escherichia coli* the cells become sensitive to the addition of chloroform, consistent with the action of a lytic enzyme. There is no direct evidence for a lytic role of Ms6 *ORF3* but it shares sequence similarity with other mycobacteriophage proteins that are implicated in this function (see below). The assignment of a putative holin function to the Ms6 *hol* gene is supported by sequence similarity with the putative holin of lactococcal phage r1t and the ability of Ms6 *hol* to complement a λ *S* mutant (25).

Lysis genes can be identified in most of the 14 sequenced mycobacteriophage genomes (figure 38-4). The lysis genes of phage Che8—*32*, *33*, and *34*—are organized similarly to those in Ms6, corresponding to the *lysA*, *ORF3*, and *hol* functions respectively; Che8 gp32 and gp33 share 57% and 88% amino acid sequence identity with Ms6 LysA and ORF3 respectively, and Che8 gp34 and Ms6 Hol are 98% identical over the first 66 of the 77 residues. However, phage Che8 is the only one that shares this organization with phage Ms6 (figure 38-4).

All of the mycobacteriophages contain a *lysA*-like gene (figure 38-4). However, the sequence relationships among the gene products are complex and frequently only a small segment of protein pairs is related. For example, regions of the 424 residue Che8 gp32 (which is similar to Ms6 LysA over its entire length) are related to other mycobacteriophage proteins as follows: 1–146 matches TM4 gp29 (47%), 61–236 matches Bxz1 gp236 (27%), 393–425 matches Corndog gp69 (78%), and 337–421 matches Che9d gp35 (40%). There are no significant matches to other LysA-family members. A central segment of TM4 gp29 matches parts of Barnyard gp39, Bxz1 gp236, and Corndog gp69 as well as the *M. tuberculosis* protein, Rv3594. This part of these proteins corresponds to a domain

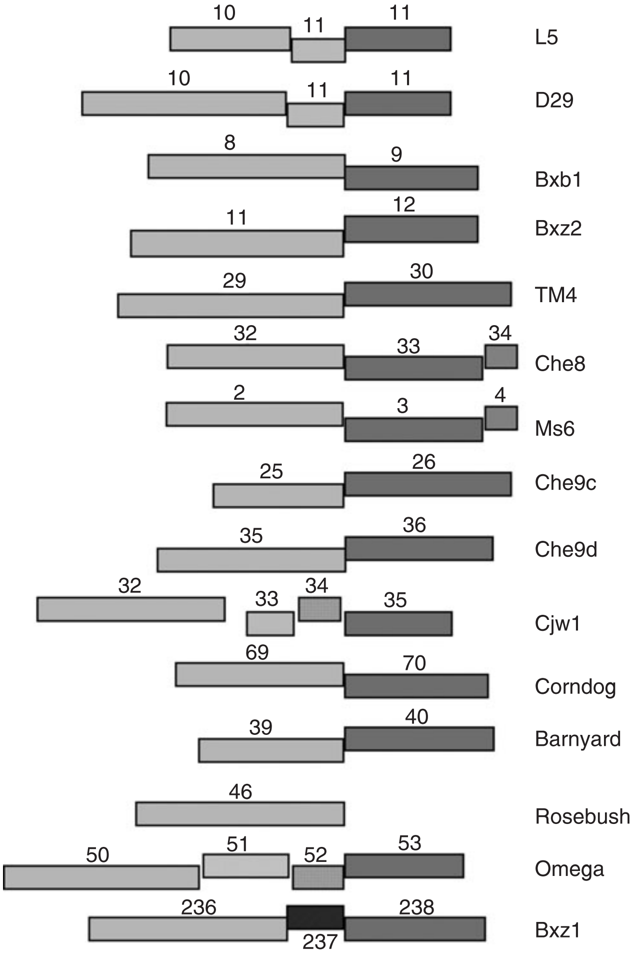


Figure 38-4 Mycobacteriophage lysis genes. With the exception of phage Rosebush, all the sequenced mycobacteriophage genomes along with the partially sequenced Ms6 genome contain two genes encoding putative lytic enzymes used for degradation of the cell wall at the completion of lytic growth. The putative *lysA* genes (light gray boxes) form a family of related sequences but the relationships are complex and not all pairs of genes have detectable sequence similarity. The *lysB* genes (dark gray boxes) are also implicated in lysis. Putative holin genes (diagonally striped boxes) can be identified in six of the phages, with L5 *11*, D29 *11*, and Cjw1 *33* forming one set of related genes, Che8 *34* and Ms6 *4* another, and Bxz1 gp237 a third.

associated with *N*-acetylmuramoyl-L-alanine amidases, consistent with a function in cell lysis. Cjw1 gp32 has a central segment related to *M. tuberculosis* Rv3766 (which is of unknown function but shares a small region of similarity with Rv3594) and has an N-terminal segment similar to Omega gp50, which in turn has a central segment with a 1,4-*β*-*N*-acetylmuramidase domain that is shared by many lysozymes. Preliminary data indicate that D29 gp10, Che8 gp32, and Bxz2 gp236 function as lytic enzymes (T. Huang, L. Marinelli, and G. F. Hatfull, unpublished observations). This LysA family of proteins thus appears to be a particularly diverse and interesting group of lytic enzymes that warrant considerable further investigation.

Phage Ms6 gp3, which lies between the *lysA* and *hol* genes, may also be involved in host lysis (25). Interestingly, homologs of this protein are found adjacent—or very close — to *lysA* in all the other mycobacteriophages; the one exception is phage Rosebush, which does not possess a homologu (figure 38-4). In general, these proteins each contain a segment of up to 250 amino acid residues at the N-terminus that are related to each other; the C-termini are more varied among the group. Some of these proteins (TM4 gp30) have a peptidoglycan-binding motif at their extreme N-terminus consistent with a role in lysis. Furthermore, preliminary data suggest that at least one of this class of proteins (Che8 gp33) exhibits lytic properties (T. Huang and G.F. Hatfull, unpublished observations). We therefore propose that these are designated as *lysB* genes.

Putative holin genes can be identified in six of the mycobacteriophages (L5, D29, Che8, Ms6, Cjw1, and Bxz1) where they are adjacent or close to the other lysin genes. Ms6 gp4 (*hol*) and its close relative, Che8 gp34, are putative class II holins (25). L5 and D29 both have a gene (gene *11*) located between the two other putative lysin genes, *10* and *12*, and preliminary data support the function of gene *11* as a holin (Marinelli and G.F. Hatfull, unpublished observations); Cjw1 gp33 is a homolog of L5 and D29 gp11. Preliminary data also support the function of the phage Bxz1 gene *237* as a holin (Huang and G.F. Hatfull, unpublished observations). Holin genes have yet to be identified in the other mycobacteriophages.

Genome Evolution

The 14 completely sequenced mycobacteriophage genomes are a highly varied group. What are their evolutionary histories and what mechanisms gave rise to their present structures? The availability of a group of genomes for comparative analysis provides an opportunity to address these questions, but with such a highly diverse group—even with nearly 1 Mbp of total sequence information and over 1600 genes—that the dataset is still inadequate for any detailed reconstruction of evolutionary histories.

In spite of these limitations, there are some striking features that reveal the dominant processes in bacteriophage evolution. Perhaps the most obvious of these is the pervasive mosaicism, with each genome being composed of modules that are shared by one or more of the other phages. The "sharing" as we define it here for the most part means having significant sequence similarity of protein products since, with some notable exceptions, most of the genomes share little or no nucleotide sequence similarity. The evolutionary events (i.e., changes at the genomic level) that gave rise to the present structures likely occurred at times far back in evolutionary time. What is most striking is that adjacent genomic modules (often corresponding to single genes) match modules in different mycobacteriophages. As a consequence, each module has its own evolutionary history as illustrated in figure 38-5, and different modules have different phylogenetic relationships. There is, therefore, no single phylogenetic description for the phage as a unit of evolutionary change and any attempt to blend

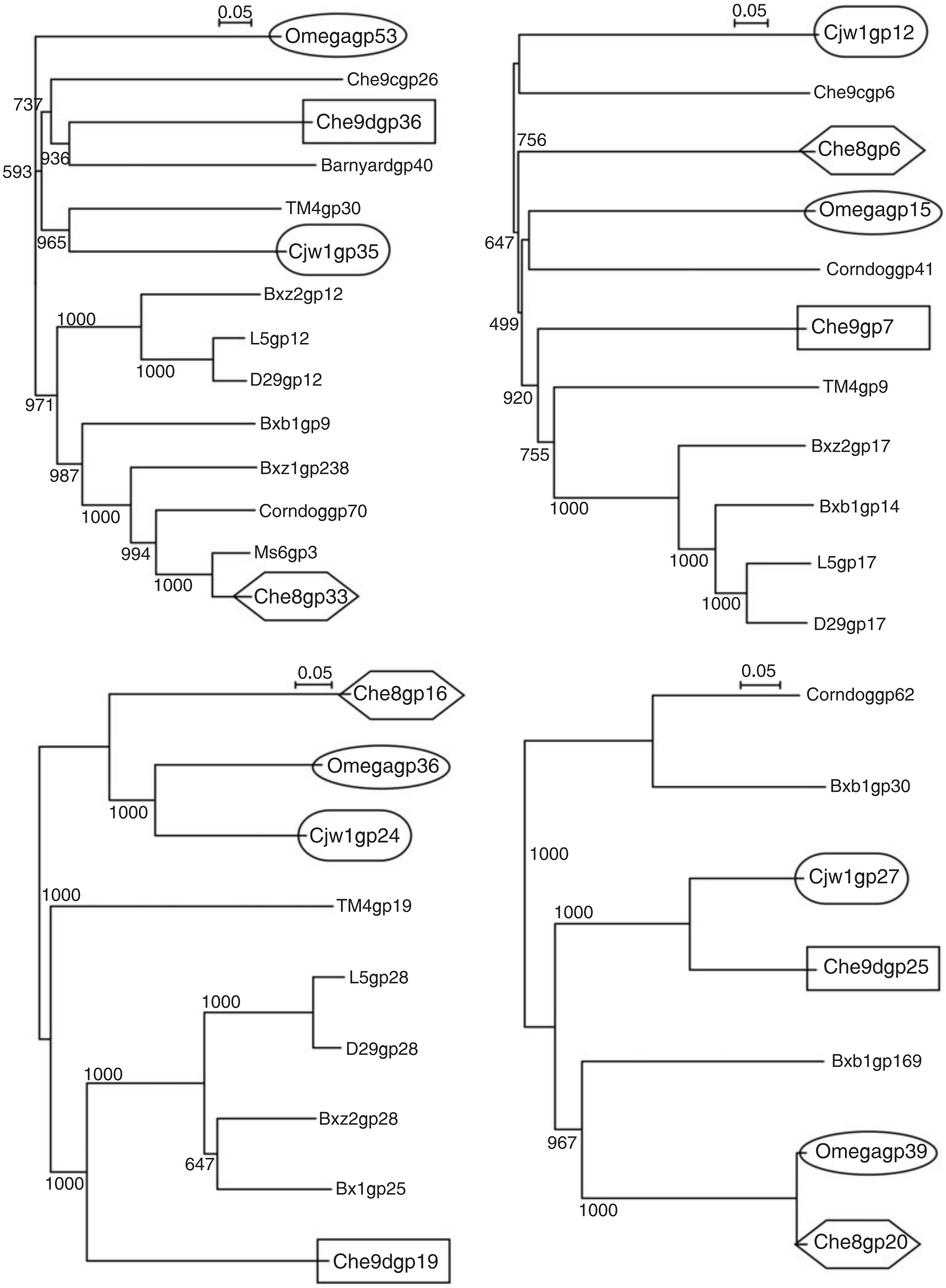

0.05
Omegagp53
Che9cgp26
Che9dgp36
Barnyardgp40
TM4gp30
Cjw1gp35
Bxz2gp12
L5gp12
D29gp12
Bxb1gp9
Bxz1gp238
Corndoggp70
Ms6gp3
Che8gp33
737
936
593
965
1000
1000
971
987
1000
994
1000
0.05
Cjw1gp12
Che9cgp6
Che8gp6
Omegagp15
Corndoggp41
Che9gp7
TM4gp9
Bxz2gp17
Bxb1gp14
L5gp17
D29gp17
756
647
499
920
755
1000
1000
1000
0.05
Che8gp16
Omegagp36
Cjw1gp24
TM4gp19
L5gp28
D29gp28
Bxz2gp28
Bx1gp25
Che9dgp19
1000
1000
1000
1000
647
1000
0.05
Corndoggp62
Bxb1gp30
Cjw1gp27
Che9dgp25
Bxb1gp169
Omegagp39
Che8gp20
1000
1000
967
1000

Figure 38-5 Phylogenetic analysis of mycobacteriophage genes. The mycobacteriophage genomes are highly mosaic and it is not possible to draw aggregate phylogenetic relationships for the phages as whole genomic units that accurately reflect their evolutionary history. Panels A–D show the phylogenetic relationships of four genes that are present in at least seven of the genomes: A: *lysB*, B: capsid subunits, C: a putative tail protein, and D: a putative D-ala-D-ala carboxypeptidase. The genes of phages Omega, Cjw1, Che8, and Che9d (shown in boxes) clearly have different evolutionary histories. Trees were generated using the neighbor-joining method in ClustalX and displayed usingNJPlot. The results of bootstrap analysis with 1000 reiterations are shown.

the many different histories into a single relationship discards the inferred histories of individual modules (44).

How do phage genomes become so pervasively mosaic? The most obvious conclusion is that the modules are exchanging by horizontal transfer from one genome to another in a way that does not require large segments of DNA homology. This is not to deny that phage replication continually generates mutations that are inherited by vertical transmission from parent to progeny, or that these mutations are re-sorted by homologous recombination—as they surely are—but these processes are not responsible for generating genomic mosaicism (35). Moreover, mosaicism is so prevalent that the mechanisms that give rise to it are anything but a minor contribution to the evolutionary process.

Typically, mosaic boundaries are found at or close to the ends of protein coding genes. In the lambdoid genomes the modules frequently consist of blocks of structural genes (37, 44) (chapter 27), although in the mycobacteriophages it is most evident outside the structural genes and modules typically are single genes. There are two plausible methods by which mosaic boundaries could be generated. First, exchange could occur at short conserved "boundary" sequences as first proposed by Susskind and Botstein, resulting in a recombinant configuration of the differing flanking genes (78). Secondly, recombination could be truly illegitimate and occur randomly with respect to nucleotide sequence information. In this case, most events would generate genomes that are inappropriately sized for packaging and likely interrupt genes required for viral propagation (11). Thus, only a minority of the genomic trash will generate viable progeny, with surviving exchange events perhaps typically occurring at or close to gene boundaries.

While short conserved "boundary" sequences have been reported in some phages of *E. coli* (12, 67), these are not generally present in the mycobacteriophages in spite of the pervasive mosaicism of these phages. Illegitimate recombination therefore seems a more likely mechanism, although acquiring supportive evidence is difficult since it is necessary to find events that have occurred recently, where the sites of recombination can be determined. Fortunately, comparative analysis of the mycobacteriophage genomes reveals several instances where two phages share almost identical DNA sequences, and there is one segment of 378 bp in common between Corndog and Che8 that is 100% identical (57). The recombination events involved did not occur at gene boundaries, but rather within protein coding sequences—albeit not far from the ends of genes—without any evidence for sequence similarity.

Who participates in these events? It is unlikely that these exchange events occur predominantly between two coinfecting phages since this is not a high-probability scenario in most natural environments. It seems more likely that recombination occurs between infecting phages and resident prophages. There are two important consequences of this supposition. First, in events where a resident prophage acquires segments of DNA, there is no immediate concern with size of the prophage. This may then participate in subsequent events that are temporally separate from the first and provide a second chance to generate a genomic structure that can give rise to viable viral particles. Secondly, since these recombination events are occurring illegitimately, then recombination can occur between an infecting phage genome and any part of the resident bacterial chromosome. This accounts for the frequent presence of genes in the mycobacteriophage genomes that have been previously thought of as bacterial genes and supports the view that bacteriophages play a major role in horizontal genetic exchange between bacterial species.

Gene Expression and Regulation

The pattern of gene expression has been most closely investigated in the temperate mycobacteriophage L5 (9, 15, 29, 34, 56). During lytic growth, two general patterns of expression are observed: synthesis of the right-arm genes early in the cycle followed by late expression of left-arm (i.e., structural) genes (see figure 38-2). A promoter responsible for early expression, P_{left}, is located at the right end of the genome and promotes leftward transcription early in lytic growth but is then downregulated late in the cycle (56). The promoter responsible for late expression, and the mechanism of activation, have yet to be described.

Phage L5 lysogeny is maintained through the action of the repressor, gp71 (15). L5 gp71 is expressed in prophages by three upstream promoters and represses the P_{left} promoter through binding to a 13 bp operator (56). Curiously, the L5 genome contains over 30 similar DNA sites to which L5 gp71 binds, nearly all of which are located within short intergenic spaces or overlapping gene starts and stops (9). The 13 bp consensus sequence is well conserved and clearly asymmetric, and the sites are in the same relative orientation with respect to transcription. Reporter gene studies indicate that the binding of gp71 to these DNA sites prevents the passage of transcriptional complexes initiating from an upstream heterologous promoter (9). These binding sites are referred to as "stoperator" sites since they appear to stop transcriptional elongation and may play a role in ensuring tight downregulation of phage genes in lysogeny that would otherwise prove deleterious to bacterial growth. In the lambdoid phages a similar function may be provided by the many stem-loop terminators present throughout their genomes (chapter 9). Mycobacteriophage genomes contain few such terminators, and when present they are usually at the ends of operons (34). Phage Bxb1 has a similar regulatory scheme to L5 and the heteroimmunity of the two phages can be accounted for by the relative binding specificities of the repressors for their cognate operator and stoperator sites (39). The genome of Bxz2 appears to contain a similar regulatory system.

Two promoter elements have been identified in phage Ms6, located upstream of the lysis genes (25). Both are σ70-like promoters and are recognized by the host RNA polymerase. It is not clear how expression of the lysis genes is regulated, but the leader sequence between the transcription start sites and the first open reading frame can form two stem-loop RNA structures, one of which functions as terminator. The regulatory schemes in other mycobacteriophages have yet to be explored although WhiB-like transcription regulators in phages TM4, Che8, Che9d, Omega, and Cjw1 are worthy of investigation.

Integration and Excision

Of the 14 sequenced mycobacteriophages, nine (L5, D29, Bxb1, Bxz2, Che8, Che9c, Che9d, Cjw1, and Omega) encode putative integration systems. These systems have been shown to be active in L5 (46), D29 (22, 69), and Bxb1 (A.I Kim and G.F Hatfull, unpublished observations). Active systems have also been demonstrated in Ms6 (24), FRAT1 (26), and the *M. tuberculosis* prophage-like element ϕRv1 (7). The integrases encoded by L5, D29, Che8, Che9c, Che9d, Cjw1, Omega, Ms6, and FRAT1 are all of the tyrosine family of recombinases, whereas those of phages Bxb1, Bxz2, and ϕRv1 are serine recombinases. Several of these systems have been adapted for the development of genetic tools (see below), and phages L5, FRAT1, and Ms6 have all been shown to integrate into tRNA genes in *M. smegmatis* or *M. tuberculosis* (24, 26, 46).

Integration and Excision of L5

The general scheme for integration and excision of L5 is similar to that of phage λ (2). The L5-encoded integrase (L5-Int) catalyzes site-specific recombination between the phage *attP* site and the chromosomal *attB* site to promote integration, but requires the host-encoded mycobacterial integration host factor, mIHF (figure 38-6) (45, 59). Integration generates two new attachment site junctions, *attL* and *attR*, and these act as substrates for excisive recombination, which is catalyzed by L5-Int and requires both mIHF and a second phage-encoded protein, L5-Xis, the product of gene *36* (48) (see chapter 7 for an overview of prophage integration and excision).

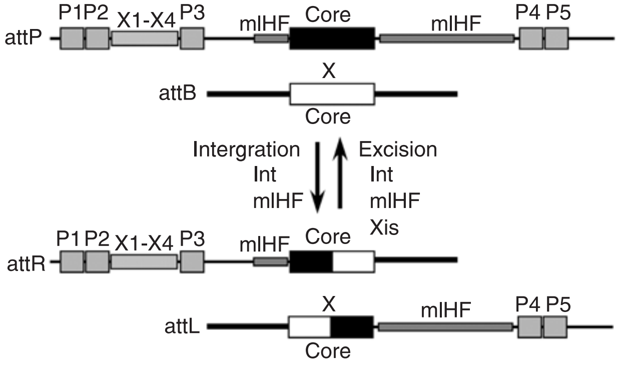


Figure 38-6 DNA sites for L5 integration and excision. The L5 *attP* and *M. smegmatis attB* site each contain a 43bp common core which in *attP* is flanked by arm-type integrase-binding sites (P1, P2, P3, P4, and P5); two additional sites (P6 and P7, not shown) to the right of P5 — as well as P3 — are dispensable for both integration and excision. The mIHF host factor does not bind specifically to *att* site DNA but occupies the indicated areas in the presence of integrase. Integration requires both integrase (Int) and mIHF); excision requires Xis in addition to Int and mIHF (see chapter 7 for a general discussion of these mechanisms. Xis binds specifically to four closely spaced sites (X1–X4) adjacent to P2.

The *attP* and *attB* sites for L5 integration share a 43 bp identical sequence (the common core) within which strand exchange occurs (46); L5-Int cleaves on either side of a 7 bp overlap region at the left end of the core (64). The functional *attP* site encompasses approximately 240 bp of DNA which contains two types of binding sites: core-type sites on each side of the sites of strand exchange and arm-type sites that flank the core (figure 38-6) (63). There are seven arm-type sites, but only four—two pairs of closely spaced sites—are needed for either integration or excision (figure 38-6).

L5-Int is a far-distant relative of λ integrase at the sequence level, but has a similar structural organization, with a small N-terminal domain that binds to arm-type sites and a large catalytic domain that binds to core-type binding sites (45, 63). L5-Int therefore has the potential to form intra- and intermolecular bridges with the two domains bound simultaneously to core- and arm-type sites. The role of mIHF, which does not bind by itself site-specifically to *attP* DNA, appears to promote intramolecular protein-bridges by facilitating or stabilizing DNA bends (58).

Integration can proceed through the DNA binding of phage L5-Int protein and host-factor mIHF to form an intasome complex that is then able to capture *attB* DNA through intermolecular Int bridges between the P1/P2 arm-type Int binding sites in *attP*, and *attB* (60, 61). This process illustrates an interesting feature of L5-Int, in that in the absence of *attB* DNA, the P1/P2 pair of sites is unbound by L5-Int, even though L5-Int is present. When *attB* DNA is added, the intermolecular bridge is formed suggesting, perhaps, that the binding of Int to the P1/P2 arm-type sites is stimulated when the core domain is also occupied (60). This interaction can occur in the absence of intasome formation, suggesting an alternative assembly pathway for recombination (60).

When both DNAs and both proteins are present, a synaptic complex is formed that is a direct precursor of a presumed recombinagenic complex (60). Recombination by this synaptic complex is stimulated by mIHF and by DNA supercoiling, which can be provided by either the *attP* or *attB* substrate (62). When the reaction products are formed, *attL* is released as an Intasome-L complex, whereas *attR* is observed by native gel electrophoresis as unbound DNA (60); however, DNA footprinting shows that the L5-Int protein remains bound to the core sites of *attR* with the mIHF protein also present and bound adjacent and to the left of L5-Int (Lewis and Hatfull, unpublished observations).

The recombination directionality factor, L5-Xis, not only stimulates excision but also inhibits integration (J.A. Lewis and G.F. Hatfull, unpublished observations). L5-Xis acts by binding to four sites (X1-X4) located between the core and the P1/P2 arm-type sites and introducing a DNA bend (figure 38-6). This bend facilitates the formation of an intasome complex with *attR* DNA that contains intramolecular bridges between the core and P1/P2 arm-type sites (J.A. Lewis and G.F. Hatfull, unpublished observations). The formation of this Intasome-R complex appears to be necessary for productive synapsis with an Intasome-L complex, although the reason is not yet clear. L5-Xis actively inhibits integration by preventing the conversion of synaptic complex 1 to the presumed recombinagenic complex within which strand exchange occurs (J.A. Lewis and G.F. Hatfull, unpublished observations).

Integration and Excision of ϕRv1

The prophage-like element, ϕRv1, is a resident of both sequenced strains of *M. tuberculosis* (13, 21) but is absent from the vaccine strain *M. bovis* BCG (50). It encompasses approximately 10 kbp and is bordered by two 12 bp direct repeats—of which there is only a single copy in BCG—and encodes a protein of the serine recombinase family of site-specific recombinases (75). Curiously, the element is situated at two different chromosomal locations in the two sequenced *M. tuberculosis* strains, suggesting that although it is unlikely to produce infectious particles it nevertheless is mobile. The recombination system has been shown to be active and utilizes an *attB* site that is part the REP13E12 repeat of which there are seven divergent copies in *M. tuberculosis* (7). The ϕRv1 element can integrate into four of these repeats and multiple integration events can be observed. A ϕRv1 gene encoding a recombination directionality factor has also been identified although the mechanism by which it influences the directionality of recombination remains to be elucidated (7).

Applications and Biotechnology

Mycobacteriophages represent wonderful toolboxes for the development of mycobacterial genetics and novel methods for the control of mycobacterial diseases (30). While their full potential has yet to be realized, several useful advances have been made which will be briefly discussed here.

Vector Development

The integration systems of L5, FRAT1, and Ms6 have all been used to construct integration-proficient plasmid vectors. In each case, a cassette containing the *attP* site and the integrase gene is inserted into a nonreplicating plasmid (which also contains an *E. coli* origin of replication) (24, 27, 77). These integration vectors typically transform both fast- and slow-growing mycobacteria efficiently to generate recombinant strains with a single integrated copy of the plasmid inserted at the *attB* site. The advantages of these vectors are that they are maintained more stably than extrachromosomal plasmid vectors (46) and are present as a single copy, an important feature when phenotypic effects resulting from multicopy plasmids are observed (6). Since integrase can mediate excisive recombination at low frequency in the absence of a recombination directionality factor, further genetic stability can be achieved by providing the integrase gene on a nonreplicating plasmid that is subsequently lost after integration of an *attP*-containing integration vector (63).

The use of phage repressor genes provides a further aspect to vector development. L5 gene *71* encodes the phage repressor that is required for lysogenic maintenance, confers immunity to superinfection (15), and can be used as a genetically selectable marker. The potential advantage of such a feature is that it avoids the use of drug-resistant genes, which are undesirable in the development of live recombinant vaccines (15). Furthermore, there is still a rather limited repertoire of positive selectable markers for mycobacterial genetics and the application of a variety of heteroimmune phage immunity systems could help to alleviate this problem.

Transduction

Two mycobacteriophages have been reported to be capable of generalized transduction: I3 (66) and Bxz1 (W.R. Jacobs and G.F. Hatfull, unpublished observations). Both these phages infect *M. smegmatis* and transfer genetic markers at varied frequencies. No phages have been described for general transduction of *M. tuberculosis* and this remains an important limitation on mycobacterial genetic systems.

The use of recombinant mycobacteriophages for specialized transduction provides a useful method for gene replacement or mutagenesis in *M. tuberculosis* (4). Targeted gene replacement can be problematical in *M. tuberculosis* since simple electroporation with nonreplicating plasmids, followed by selection, yields a high proportion of illegitimate recombination events, even when chromosomal DNA is present on the plasmid molecule (41). Specialized transducing phages can be readily constructed by cloning

Polymerase chain reaction fragments of sequence that normally flanks the to-be-replaced gene into a cosmid vector such that these flanking sequences now flank a drug-resistance gene. This construct is then inserted into a conditionally replicating TM4-derived shuttle phasmid and recovered in *E.* coli via in vitro packaging into λ particles. DNA of the new recombinant phasmids is then electroporated into *M. smegmatis* to generate infectious phage particles, which can then be used to infect *M. tuberculosis* at nonpermissive temperatures. Typically, a high proportion of the bacteria selected for drug resistance contain a replacement of the gene of interest with the drug-resistance marker (4). If the replaced gene is essential then the infection can be performed in a strain into which a second copy of the gene has been introduced.

Transposon Delivery

Conditionally replicating phages are also useful as transposon-delivery systems (5). A series of shuttle phasmids have been constructed in conditionally replicating phages that contain transposons such as mini-Tn10 and Tn5367 that can be used to infect either fast- or slow-growing mycobacteria. Selection for the drug marker yields progeny in which the transposon has moved from the phage onto the bacterial chromosome. Similar phages have also been described for transposon delivery in *M. paratuberculosis* (28). The locations of the mutations that give rise to the conditionally replicating phenotype of these phages have not yet been mapped.

Use of Mycobacteriophages for Diagnosis of Tuberculosis

Two phage-based systems for the diagnosis of tuberculosis infections have been described. One is the PhaB assay, in which a mycobacteriophage is used to infect a clinical sample suspected of containing *M. tuberculosis* and the number of phage particles generated is evaluated by plating on *M. smegmatis*; the presence of *M. tuberculosis* in the sample can be inferred by an increase in the number of plaques (82). Furthermore, this method can be used to determine drug susceptibility profiles by examining the effects of antibiotics on phage production (19). Comparisons of this method with other traditional approaches for *M. tuberculosis* diagnosis and drug susceptibility testing have been conducted with encouraging results (16–19, 52, 81). A second method employs the use of recombinant mycobacteriophages carrying the luciferase reporter gene (38, 73). These FFlux reporter phages produce light when they infect *M. tuberculosis* and can be used to evaluate empirical drug susceptibility profiles by inclusion of antibiotics in the assay. This system offers good sensitivity and reliable drug susceptibility results and shows considerable promise in a clinical setting (3, 70–72). See chapter 46 for additional discussion of phage-based bacterial identification.

Concluding Remarks

Mycobacteriophages have proven themselves to be valuable toolboxes for mycobacterial genetics. However, as we begin to learn more about the genomic diversity of this group of bacteriophages it is quite evident that we know very little about the greater population of mycobacteriophages in the biosphere. The 14 completely sequenced mycobacteriophage genomes will require considerable dissection, both experimentally and bioinformatically, to understand the features that are already apparent. As the database of mycobacteriophage genomic information advances, we anticipate many new insights and expect a wealth of information that can be used to further our understanding of their pathogenic mycobacterial hosts.

References

1. Anne, J., P. Fiten, L. Van Mellaert, B. Joris, G. Opdenakker, and H. Eyssen. 1995. Analysis of the open reading frames of the main capsid proteins of actinophage VWB. Arch. Virol. 140:1033–1047.
2. Azuro, M. A. and A. Landy. 2002. λ Int and the λ Int family, pp. 118–148. *In* A. M. Lambowitz (ed.) Mobile DNA II. American Scoiety for Microbiology, Washington, D.C.
3. Banaiee, N., M. Bobadilla-Del-Valle, S. Bardarov, Jr., P. F. Riska, P. M. Small, A. Ponce-De-Leon, W. R. Jacobs, Jr., G. F. Hatfull, and J. Sifuentes-Osornio. 2001. Luciferase reporter mycobacteriophages for detection, identification, and antibiotic susceptibility testing of *Mycobacterium tuberculosis* in Mexico. J.Clin. Microbiol. 39:3883–3888.
4. Bardarov, S., S. Bardarov, Jr., M. S. Pavelka, Jr., V. Sambandamurthy, M. Larsen, J. Tufariello, J. Chan, G. Hatfull, and W. R. Jacobs, Jr. 2002. Specialized transduction: an efficient method for generating marked and unmarked targeted gene disruptions in *Mycobacterium tuberculosis, M. bovis* BCG and *M. smegmatis*. Microbiology 148:3007–3017.
5. Bardarov, S., J. Kriakov, C. Carriere, S. Yu, C. Vaamonde, R. A. McAdam, B. R. Bloom, G. F. Hatfull, and W. R. Jacobs, Jr. 1997. Conditionally replicating mycobacteriophages: a system for transposon delivery to *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. USA 94:10961–10966.
6. Barsom, E. K., and G. F. Hatfull. 1996. Characterization of *Mycobacterium smegmatis* gene that confers resistance to phages L5 and D29 when overexpressed. Mol. Microbiol. 21:159–170.
7. Bibb, L. A., and G. F. Hatfull. 2002. Integration and excision of the *Mycobacterium tuberculosis* prophage-like element, phiRv1. Mol. Microbiol. 45:1515–1526.

8. Bloom, B. R. 1994. Tuberculosis: Pathogenesis, Protection and Control. American Society to Microbiology, Washington, D.C.
9. Brown, K. L., G. J. Sarkis, C. Wadsworth, and G. F. Hatfull. 1997. Transcriptional silencing by the mycobacteriophage L5 repressor. EMBO J. 16:5914–5921.
10. Brussow, H. 2001. Phages of dairy bacteria. Annu. Rev. Microbiol. 55:283–303.
11. Brussow, H., and R. W. Hendrix. 2002. Phage genomics: small is beautiful. Cell 108:13–16.
12. Clark, A. J., W. Inwood, T. Cloutier, and T. S. Dhillon. 2001. Nucleotide sequence of coliphage HK620 and the evolution of lambdoid phages. J. Mol. Biol. 311:657–679.
13. Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry, 3rd, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, B. G. Barrell, et al. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 393:537–544.
14. Cole, S. T., K. Eiglmeier, J. Parkhill, K. D. James, N. R. Thomson, P. R. Wheeler, N. Honore, T. Garnier, C. Churcher, D. Harris, K. Mungall, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. M. Davies, K. Devlin, S. Duthoy, T. Feltwell, A. Fraser, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, C. Lacroix, J. Maclean, S. Moule, L. Murphy, K. Oliver, M. A. Quail, M. A. Rajandream, K. M. Rutherford, S. Rutter, K. Seeger, S. Simon, M. Simmonds, J. Skelton, R. Squares, S. Squares, K. Stevens, K. Taylor, S. Whitehead, J. R. Woodward, and B. G. Barrell. 2001. Massive gene decay in the leprosy bacillus. Nature 409:1007–1011.
15. Donnelly-Wu, M. K., W. R. Jacobs, Jr., and G. F. Hatfull. 1993. Superinfection immunity of mycobacteriophage L5: applications for genetic transformation of mycobacteria. Mol. Microbiol. 7:407–417.
16. Eltringham, I. J., F. A. Drobniewski, J. A. Mangan, P. D. Butcher, and S. M. Wilson. 1999. Evaluation of reverse transcription-PCR and a bacteriophage-based assay for rapid phenotypic detection of rifampin resistance in clinical isolates of *Mycobacterium tuberculosis.* J. Clin. Microbiol. 37:3524–3527.
17. Eltringham, I. J., F. A. Drobniewski, J. A. Mangan, P. D. Butcher, and S. M. Wilson. 1999. Evaluation of reverse transcription-PCR and a bacteriophage-based assay for rapid phenotypic detection of rifampin resistance in clinical isolates of *mycobacterium tuberculosis* J. Clin. Microbiol. 37:3524–3527.
18. Eltringham, I. J., S. M. Wilson, and F. A. Drobniewski. 1999. Evaluation of a bacteriophage-based assay (phage amplified biologically assay) as a rapid screen for resistance to isoniazid, ethambutol, streptomycin, pyrazinamide, and ciprofloxacin among clinical isolates of *Mycobacterium tuberculosis*. J. Clin. Microbiol. 37:3528–3532.
19. Eltringham, I. J., S. M. Wilson, and F. A. Drobniewski. 1999. Evaluation of a bacteriophage-based assay (phage amplified biologically assay) as a rapid screen for resistance to isoniazid, ethambutol, streptomycin, pyrazinamide, and ciprofloxacin among clinical isolates of *Mycobacterium tuberculosis*. J. Clin. Microbiol. 37:3528–3532.
20. Fischetti, V. A. 2001. Phage antibacterials make a comeback. Nat. Biotechnol. 19:734–735.
21. Fleischmann, R. D., D. Alland, J. A. Eisen, L. Carpenter, O. White, J. Peterson, R. DeBoy, R. Dodson, M. Gwinn, D. Haft, E. Hickey, J. F. Kolonay, W. C. Nelson, L. A. Umayam, M. Ermolaeva, S. L. Salzberg, A. Delcher, T. Utterback, J. Weidman, H. Khouri, J. Gill, A. Mikula, W. Bishai, W. R. Jacobs, Jr., J. C. Venter, and C. M. Fraser. 2002. Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. J. Bacteriol. 184:5479–5490.
22. Ford, M. E., G. J. Sarkis, A. E. Belanger, R. W. Hendrix, and G. F. Hatfull. 1998. Genome structure of mycobacteriophage D29: implications for phage evolution. J. Mol. Biol. 279:143–164.
23. Ford, M. E., C. Stenstrom, R. W. Hendrix, and G. F. Hatfull. 1998. Mycobacteriophage TM4: genome structure and gene expression. Tuber. Lung Dis. 79:63–73.
24. Freitas-Vieira, A., E. Anes, and J. Moniz-Pereira. 1998. The site-specific recombination locus of mycobacteriophage Ms6 determines DNA integration at the tRNA(Ala) gene of *Mycobacterium* spp. Microbiology 144:3397–3406.
25. Garcia, M., M. Pimentel, and J. Moniz-Pereira. 2002. Expression of mycobacteriophage Ms6 lysis genes is driven by two sigma(70)-like promoters and is dependent on a transcription termination signal present in the leader RNA. J. Bacteriol. 184:3034–3043.
26. Haeseleer, F., J. F. Pollet, A. Bollen, and P. Jacobs. 1992. Molecular cloning and sequencing of the attachment site and integrase gene of the temperate mycobacteriophage FRAT1. Nucleic Acids Res. 20:1420.
27. Haeseleer, F., J. F. Pollet, M. Haumont, A. Bollen, and P. Jacobs. 1993. Stable integration and expression of the *Plasmodium falciparum* circumsporozoite protein coding sequence in mycobacteria. Mol. Biochem. Parasitol. 57:117–126.
28. Harris, N. B., Z. Feng, X. Liu, S. L. Cirillo, J. D. Cirillo, and R. G. Barletta. 1999. Development of a transposon mutagenesis system for *Mycobacterium avium* subsp. *paratuberculosis*. FEMS Microbiol. Lett. 175:21–26.
29. Hatfull, G. F. 2000. Molecular genetics of mycobacteriophages, pp. 37–54. *In* G. F. Hatfull and W. R. Jacobs, Jr. (eds.) Molecular Genetics of the Mycobacteria. ASM Press, Washington, D.C.
30. Hatfull, G. F. 1994. Mycobacteriophage L5: a toolbox for tuberculosis. ASM News 60:255–260.
31. Hatfull, G. F. 1999. Mycobacteriophages, pp. 38–58. *In* C. Ratledge and J. Dale (eds.) Mycobacteria: Molecular Biology and Virulence. Chapman and Hall, London.
32. Hatfull, G. F., and W. R. Jacobs, Jr. 2000. Molecular Genetics of the Mycobacteria. ASM Press, Washington, D.C.
33. Hatfull, G. F., and W. R. Jacobs, Jr. 1994. Mycobacteriophages: cornerstones of mycobacterial research, pp. 165–183. *In* B. R. Bloom (ed.) Tuberculosis: Pathogenesis, Protection and Control. ASM Press, Washington, D.C.
34. Hatfull, G. F., and G. J. Sarkis. 1993. DNA sequence, structure and gene expression of mycobacteriophage

L5: a phage system for mycobacterial genetics. Mol. Microbiol. 7:395–405.

35. Hendrix, R. W. 2002. Bacteriophages: evolution of the majority. Theor. Pop. Biol. 61:471–480.
36. Hendrix, R. W., J. G. Lawrence, G. F. Hatfull, and S. Casjens. 2000. The origins and ongoing evolution of viruses. Trends Microbiol. 8:504–508.
37. Hendrix, R. W., M. C. Smith, R. N. Burns, M. E. Ford, and G. F. Hatfull. 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. Proc. Natl. Acad. Sci. USA 96:2192–2197.
38. Jacobs, W. R., Jr., R. G. Barletta, R. Udani, J. Chan, G. Kalkut, G. Sosne, T. Kieser, G. J. Sarkis, G. F. Hatfull, and B. R. Bloom. 1993. Rapid assessment of drug susceptibilities of *Mycobacterium tuberculosis* by means of luciferase reporter phages. Science 260:819–822.
39. Jain, S., and G. F. Hatfull. 2000. Transcriptional regulation and immunity in mycobacteriophage Bxb1. Mol. Microbiol. 38:971–985.
40. Juhala, R. J., M. E. Ford, R. L. Duda, A. Youlton, G. F. Hatfull, and R. W. Hendrix. 2000. Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. J. Mol. Biol. 299:27–51.
41. Kalpana, G. V., B. R. Bloom, and W. R. Jacobs, Jr. 1991. Insertional mutagenesis and illegitimate recombination in mycobacteria. Proc. Natl. Acad. Sci. USA 88:5433–5437.
42. Katsura, I., and R. W. Hendrix. 1984. Length determination in bacteriophage lambda tails. Cell 39:691–698.
43. Kunisawa, T. 2000. Functional role of mycobacteriophage transfer RNAs. J. Theor. Biol. 205:167–170.
44. Lawrence, J. G., G. F. Hatfull, and R. W. Hendrix. 2002. Imbroglios of viral taxonomy: genetic exchange and failings of phenetic approaches. J. Bacteriol. 184:4891–4905.
45. Lee, M. H., and G. F. Hatfull. 1993. Mycobacteriophage L5 integrase-mediated site-specific integration in vitro. J. Bacteriol. 175:6836–6841.
46. Lee, M. H., L. Pascopella, W. R. Jacobs, Jr., and G. F. Hatfull. 1991. Site-specific integration of mycobacteriophage L5: integration-proficient vectors for *Mycobacterium smegmatis*, *Mycobacterium tuberculosis*, and bacille Calmette-Guerin. Proc. Natl. Acad. Sci. USA 88:3111–3115.
47. Levin, M. E., R. W. Hendrix, and S. R. Casjens. 1993. A programmed translational frameshift is required for the synthesis of a bacteriophage lambda tail assembly protein. J. Mol. Biol. 234:124–139.
48. Lewis, J. A., and G. F. Hatfull. 2000. Identification and characterization of mycobacteriophage L5 excisionase. Mol. Microbiol. 35:350–360.
49. Loeffler, J. M., D. Nelson, and V. A. Fischetti. 2001. Rapid killing of *Streptococcus pneumoniae* with a bacteriophage cell wall hydrolase. Science 294:2170–2172.
50. Mahairas, G. G., P. J. Sabo, M. J. Hickey, D. C. Singh, and C. K. Stover. 1996. Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis.* J. Bacteriol. 178:1274–1282.
51. McNerney, R. 1999. TB: the return of the phage. A review of fifty years of mycobacteriophage research. Int. J. Tuberc. Lung. Dis. 3:179–184.
52. McNerney, R., P. Kiepiela, K. S. Bishop, P. M. Nye, and N. G. Stoker. 2000. Rapid screening of *Mycobacterium tuberculosis* for susceptibility to rifampicin and streptomycin. Int. J. Tuberc. Lung Dis. 4:69–75.
53. Mediavilla, J., S. Jain, J. Kriakov, M. E. Ford, R. L. Duda, W. R. Jacobs, Jr., R. W. Hendrix, and G. F. Hatfull. 2000. Genome organization and characterization of mycobacteriophage Bxb1. Mol. Microbiol. 38:955–970.
54. Morgan, G. J., G. F. Hatfull, S. Casjens, and R. W. Hendrix. 2002. Bacteriophage Mu genome sequence: analysis and comparison with Mu-like prophages in *Haemophilus, Neisseria and Deinococcus*. J. Mol. Biol. 317:337–359.
55. Mukamolova, G. V., A. S. Kaprelyants, D. I. Young, M. Young, and D. B. Kell. 1998. A bacterial cytokine. Proc. Natl. Acad. Sci. USA 95:8916–8921.
56. Nesbit, C. E., M. E. Levin, M. K. Donnelly-Wu, and G. F. Hatfull. 1995. Transcriptional regulation of repressor synthesis in mycobacteriophage L5. Mol. Microbiol. 17:1045–1056.
57. Pedulla, M. L., M. E. Ford, J. M. Houtz, T. Karthikeyan, C. Wadsworth, J. A. Lewis, D. Jacobs-Sera, J. Falbo, J. Gross, N. R. Pannunzio, W. Brucker, V. Kumar, J. Kandasamy, L. Keenan, S. Bardarov, J. Kriakov, J. G. Lawrence, W. R. Jacobs, R. W. Hendrix, and G. F. Hatfull. 2003. Origins of highly mosaic mycobacteriophage genomes. Cell 113:171–182.
58. Pedulla, M. L., and G. F. Hatfull. 1998. Characterization of the mIHF gene of *Mycobacterium smegmatis.* J. Bacteriol. 180:5473–5477.
59. Pedulla, M. L., M. H. Lee, D. C. Lever, and G. F. Hatfull. 1996. A novel host factor for integration of mycobacteriophage L5. Proc. Natl. Acad. Sci. USA 93:15411–15416.
60. Pena, C. E., J. M. Kahlenberg, and G. F. Hatfull. 2000. Assembly and activation of site-specific recombination complexes. Proc. Natl. Acad. Sci. USA 97:7760–7765.
61. Pena, C. E., J. M. Kahlenberg, and G. F. Hatfull. 1999. Protein–DNA complexes in mycobacteriophage L5 integrative recombination. J. Bacteriol. 181:454–461.
62. Pena, C. E., J. M. Kahlenberg, and G. F. Hatfull. 1998. The role of supercoiling in mycobacteriophage L5 integrative recombination. Nucleic Acids Res. 26:4012–4018.
63. Peña, C. E., M. H. Lee, M. L. Pedulla, and G. F. Hatfull. 1997. Characterization of the mycobacteriophage L5 attachment site, attP. J. Mol. Biol. 266:76–92.
64. Peña, C. E., J. E. Stoner, and G. F. Hatfull. 1996. Positions of strand exchange in mycobacteriophage L5 integration and characterization of the attB site. J. Bacteriol. 178:5533–5536.
65. Popa, M. P., T. A. McKelvey, J. Hempel, and R. W. Hendrix. 1991. Bacteriophage HK97 structure: wholesale covalent cross-linking between the major head shell subunits. J. Virol. 65:3227–3237.
66. Raj, C. V., and T. Ramakrishnan. 1970. Transduction in *Mycobacterium smegmatis.* Nature 228:280–281.
67. Recktenwald, J., and H. Schmidt. 2002. The nucleotide sequence of Shiga toxin (Stx) 2e-encoding phage phiP27 is not related to other Stx phage genomes, but the modular genetic structure is conserved. Infect. Immun. 70:1896–1908.

68. Redmond, W. B., and D. M. Ward. 1966. Media and methods for phage-typing mycobacteria. Bull. WHO 35:563–568.
69. Ribeiro, G., M. Viveiros, H. L. David, and J. V. Costa. 1997. Mycobacteriophage D29 contains an integration system similar to that of the temperate mycobacteriophage L5. Microbiology 143:2701–2708.
70. Riska, P. F., and W. R. Jacobs, Jr. 1998. The use of luciferase-reporter phage for antibiotic-susceptibility testing of mycobacteria. Methods Mol. Biol. 101:431–455.
71. Riska, P. F., W. R. Jacobs, Jr., B. R. Bloom, J. McKitrick, and J. Chan. 1997. Specific identification of *Mycobacterium tuberculosis* with the luciferase reporter mycobacteriophage: use of *p*-nitro-alpha-acetylamino-beta-hydroxy propiophenone. J. Clin. Microbiol. 35:3225–3231.
72. Riska, P. F., Y. Su, S. Bardarov, L. Freundlich, G. Sarkis, G. Hatfull, C. Carriere, V. Kumar, J. Chan, and W. R. Jacobs, Jr. 1999. Rapid film-based determination of antibiotic susceptibilities of *Mycobacterium tuberculosis* strains by using a luciferase reporter phage and the Bronx Box. J. Clin. Microbiol. 37:1144–1149.
73. Sarkis, G. J., W. R. Jacobs, Jr., and G. F. Hatfull. 1995. L5 luciferase reporter mycobacteriophages: a sensitive tool for the detection and assay of live mycobacteria. Mol. Microbiol. 15:1055–1067.
74. Schuch, R., D. Nelson, and V. A. Fischetti. 2002. A bacteriolytic agent that detects and kills *Bacillus anthracis.* Nature 418:884–889.
75. Smith, M. C., and H. M. Thorpe. 2002. Diversity in the serine recombinases. Mol. Microbiol. 44:299–307.
76. Snider, D. E., Jr., W. D. Jones, and R. C. Good. 1984. The usefulness of phage typing *Mycobacterium tuberculosis* isolates. Am. Rev. Respir. Dis. 130:1095–1099.
77. Stover, C. K., V. F. de la Cruz, T. R. Fuerst, J. E. Burlein, L. A. Benson, L. T. Bennett, G. P. Bansal, J. F. Young, M. H. Lee, G. F. Hatfull, et al. 1991. New use of BCG for recombinant vaccines. Nature 351:456–460.
78. Susskind, M. M., and D. Botstein. 1978. Molecular genetics of bacteriophage P22. Microbiol. Rev. 42:385–413.
79. van Sinderen, D., H. Karsens, J. Kok, P. Terpstra, M. H. Ruiters, G. Venema, and A. Nauta. 1996. Sequence analysis and molecular characterization of the temperate lactococcal bacteriophage r1t. Mol. Microbiol. 19:1343–1355.
80. Wang, I. N., D. L. Smith, and R. Young. 2000. Holins: the protein clocks of bacteriophage infections. Annu. Rev. Microbiol. 54:799–825.
81. Watterson, S. A., S. M. Wilson, M. D. Yates, and F. A. Drobniewski. 1998. Comparison of three molecular assays for rapid detection of rifampin resistance in *Mycobacterium tuberculosis.* J. Clin. Microbiol. 36:1969–1973.
82. Wilson, S. M., Z. al-Suwaidi, R. McNerney, J. Porter, and F. Drobniewski. 1997. Evaluation of a new rapid bacteriophage-based method for the drug susceptibility testing of *Mycobacterium tuberculosis.* Nat. Med. 3:465–468.
83. Wommack, K. E., and R. R. Colwell. 2000. Virioplankton: viruses in aquatic ecosystems. Microbiol. Mol. Biol. Rev. 64:69–114.
84. Xu, J. 2000. A conserved frameshift strategy in dsDNA long tailed phages. PhD thesis, University of Pittsburgh, Pittsburgh.

39

Molecular Genetics of *Streptomyces* Phages

MARGARET C. M. SMITH

Bacteria in the genus *Streptomyces* are abundant in most soil environments (103). They are prolific producers of secondary metabolites, which often have antibiotic or other biological activity, and they have an unusual growth habit (57). A streptomycete colony grows initially as a vegetative mycelium, containing branching hyphal filaments with occasional cross-walls. As the colony matures, signals govern the production of aerial hyphae, which differentiate to form chains of spores. As spores are resistant to drought and can persist in the soil for many years, the high streptomycete count in soils is often attributable to the presence of spores (16). The isolation of phages from soil samples using a *Streptomyces* host is generally easy with most samples yielding phage after enrichment (57). *Streptomyces coelicolor* A3(2) is a model organism for this genus and has been studied most intensively. The linear genome contains 8,667,507 bp (5). Although the genome does not have any obvious prophage sequences, it does contain about 11 elements with features reminiscent of plasmids.

Studies on phage biology often lead to the development of genetic tools for the manipulation of the host. Interest in *Streptomyces* species frequently relates to antibiotic biosynthesis and to the use of genetic techniques to generate novel secondary metabolites (50, 56, 77). Thus, the ability to transfer DNA between species is highly desirable, but streptomycetes vary substantially in their susceptibility to genetic manipulation. Phage-derived tools have made a major impact on genetic engineering of *Streptomyces* and their close relatives, *Saccharopolyspora* (57). As it is impossible to consider the life cycle of a phage without considering the host to which it is adapted, fundamental studies of *Streptomyces* phages have led to a deeper understanding of both phage and host. One temperate phage, ϕC31, which infects both *S. coelicolor* A3(2) and its close relative *Streptomyces lividans*, has been studied extensively and has been exploited in the development of vectors and genetic tools. The excellent genetic studies by Lomovskaya and colleagues provided the sound basis for future work with ϕC31 and other *Streptomyces* phages (65). Two comprehensive accounts of the development and nature of cloning vectors from ϕC31 have been published (17, 57) and will not be discussed here. This chapter aims to summarize progress since the review by Lomovskaya et al. (65) on *Streptomyces* phage biology.

Diversity and Evolution of *Streptomyces* Phages

By far the most studied *Streptomyces* phage is ϕC31 and the major features of this phage are described below. However, many other actinophages have been isolated, not only with the intent to develop them into the genetic tools but also to satisfy the curiosity of researchers (table 39-1). All the phages have been isolated that infect *Streptomyces*, or the closely related genus *Saccharopolyspora*, are double-stranded DNA phages with similar morphologies, that is icosahedral heads and long, flexible tails (1). Where studied, the genomes are linear with *cos* ends or terminal redundancy and vary in size between 40 kbp and 121 kbp. The ends of two phages, FP22 and ϕA1, are not of a known type, that is they are not cohesive but do appear to be discrete (32, 41). Most of the phages that have been studied are temperate and the *att* sites of several that have been shown to integrate into the host chromosome have been mapped on the phage genome (see below). An important exception to these phages is the *Streptomyces fradiae* phage, ϕSF1, which exists as a plasmid in its prophage from (17, 22). A deletion analysis on the prophage has identified essential and inessential regions of the genome and those required for plaque formation or immunity. A strain containing the plasmid form can cause "pocks" (which are zones of growth inhibition), a property typical of transferable plasmids in *Streptomyces*.

Studies on phage genomics have indicated that double-stranded DNA phage genomes are mosaics, having frequently exchanged/acquired genes by horizontal gene transfer, and have access to a common gene pool (9, 10, 45) (see chapter 4 and 27). A signature of horizontal exchange is a sudden change in the degree of sequence similarity when comparing one genome with another and this is

Table 39-1 Properties of *Streptomyces* and *Saccharopolyspora* Phages

Phage	Isolation host	Genome size (kbp)	% G+C	DNA ends	Phage–phage relationship	Phage–host interactions	Application	References
ϕC31, ϕC43, ϕC62, ϕBT1, ϕSEA	*S. lividans* 66 Broad host range	All around 41.5	63.6	Cohesive	All in the same immunity group. Similar at the DNA sequence level, but with additional DNA in some genomes and evidence of mosaicism. ϕC31 DNA cross-hybridizes to that of TG1. TG1 and ϕA7 have similar genome organization of ϕC31 and ϕBT1	Temperate. ϕC31 antimodification in *S. albus*. Sensitive to Pgl. Inhibits host rRNA synthesis in lytic growth. Receptor probably a glycoprotein	Cloning and integration vectors; plasmid transduction	17, 21, 23, 57, 65, 86, 87, 92
TG1	*S. cattleya* Broad host range	41		Cohesive	Related to ϕC31 by DNA hybridization, cross-reactivity with antisera, and similar genome organization	Temperate	Phagemids; accommodate 8.5 kbp	33, 34
ϕA7	*S. antibioticus* Broad host range	46.7		Cohesive	Similar genome organization to ϕC31, ϕBT1, and TG1	Temperate		31, 32
ϕA2, ϕA4, ϕA9	*S. antibioticus* Broad host range	43, 49, 53		Cohesive	Cross-hybridize with each other	ϕA2 avoids *Sal*P1 restriction. ϕA4 avoids *Sac* restriction		32
ϕHAU3	*S. hygroscopicus* Broad host range	51	66	Cohesive	—	Temperate. *S. lividans* resistance to ϕHAU3 is due to Ea59-like endonuclease	Phagemids	107, 108
ϕA5, ϕA6	*S. antibioticus* Fairly broad host range	—		Terminally redundant	Cross-hybridize strongly to each other and faintly to ϕA8	Temperate		32
R4	*S. albus* J1074 (R^-M^-) Broad host range	53.7	67	Cohesive	R4 and SH10 have very similar restriction maps	Temperate. Antimodification in *S. albus*	Cloning and integration vectors. Plasmid transduction	17, 20, 59, 68, 71
SH10	*S. hygroscopicus* Broad host range	~41	68–73	Cohesive	Same immunity group as R4	Temperate. Inducible by UV		59, 60
VWB	*S. venezualae* Narrow host range	43.7	69.3	Cohesive	Genome organization similar to R4. Coat proteins distantly related to those of λ	Temperate	Cloning and integration vectors	2, 3, 100
ϕA8	*S. antibioticus* Broad host range	50	—	Cohesive	Probably homoimmune with R4	Temperate	—	32
SAt1	*S. azureus*	~37	71	—	—	Temperate	—	72
RP2, RP3	*S. rimosus* Narrow host range	64.7, 62.4	70	Cohesive	~400 bp cross-hybridization between RP2 and RP3	Temperate	Plasmid transduction, integration vectors	36, 58, 74

SC623, SC681, SC347	*S. coelicolor* (Müller)	57	68–71		SC681 and 623 are homoimmune	Temperate		80
FP43	*S. fradiae*, *S. griseofuscus* Broad host range	68		Terminally redundant	—	Temperate	Plasmid transduction	69
ϕA1	*S. antibioticus*	100		Discrete	—	—	—	32
FP22	*S. fradiae*, *S. griseofuscus* Broad host range	131	46	Discrete	Homoimmune with P23, and cross-hybridize	Temperate. Refractory to restriction		41
ϕSF1	*S. fradiae*	82		Terminally redundant		Prophage is a plasmid and can form pocks. Two prophage forms differ in their frequency of prophage induction	Generalized transducing phage	22
ϕSV1,3,9, 10–12	S. venezualae Narrow host range	45	—	Terminally redundant	SV1, 9, 11 are different immunity groups but cross-hybridize	Temperate	Generalized transducing phages	94, 101
DAH1, DAH2, DAH4–DAH6	*S. coelicolor* A3(2) Also plaque on *S. avermitilis* and *S. lividans*	93–121	—	—	—	—	Generalized transducing phages	12
JHJ3 (JHJ1, JHJ2)	*Saccharapolyspora hirsuta*. JHJ1 and JHJ2 broad host range in the *Saccharopolyspora* genus	41.1	70	Cohesive	JHJ1, JHJ2 are derivatives of JHJ3	JHJ3 temperate. JHJ1 is virulent JHJ1 and JHJ2 can form invasive plaques	Plasmid shuttle vectors	37, 38, 42
ϕC69	*Saccharopolyspora erythreae* Narrow host range	40	—	Cohesive	—	Virulent. Displays “lysis from without” in some *Saccharo polyspora* strains (e.g., NRRL2359)	—	55
121, SE-3, SE-5	*Saccharopolyspora erythreae* Narrow host range	41.9 (121) 42.2 (SE5)	57.5–62.5	Cohesive	Homoimmune and cross-hybridize strongly, except in central region. 121 cannot propagate on ϕFR113/114 lysogens	Virulent. SE-3 shows “lysis from without” with NRRL2359		78, 11, 93
ϕFR113/ ϕFR114	*Saccharopolyspora rectivergula* (formerly *Faenia rectivergula*)	43, 42		Cohesive	Homoimmune and cross-hybridize strongly, except in central region			79

frequently seen in comparisons of lambdoid genomes (54). A study of phage genomes of *S. coelicolor* (Müller) and *Saccharopolyspora* (previously *Faenia*) using restriction endonuclease mapping also indicates that mosaicism is rife in the actinophages (78, 80). Heteroduplex analysis of DNA of ϕC31 and its homoimmune relative, ϕC43 (isolated from *S. lividans* 803), showed that they are highly homologous for most of their lengths but also that ϕC43 has an insertion, believed to be an IS element, and a short tract of nonhomology in a region toward the center of the genome (65, 86, 87). More recently, a comparison between the DNA sequences of the complete genomes of ϕC31 and a homoimmune phage, ϕBT1, also indicated sudden changes in sequences similarity between adjacent genes from 81% to <17% similarity at the amino acid level (39). There are also segments of DNA that are present in one genome with no comparable DNA in the other and, as discussed below, these segments encode additional genes (39).

Because of frequent horizontal exchange of genes, the relationship between phage that infect the same genus is frequently obscure; indeed, the closest relatives to some ϕC31- or ϕBT1-encoded proteins in the database appear to be in phages that infect hosts evolutionarily very diverged from *Streptomyces* (30, 45, 92). For instance, there is little or no significant sequence similarity between the genes encoding the coat proteins of the actinophages ϕC31 and VWB (2, 92). In fact, ϕC31 head-assembly proteins are similar to those of a host of non-streptomycete phages including coliphage HK97, whereas the coat proteins from VWB are related to phage λ coat proteins. These observations support the idea that among streptomycete phages there are at least two lineages of head protein gene clusters (9).

Another feature of phage evolution is the maintenance of genome organization within families of phage. With the lambdoid family this facilitates homologous recombination and leads to the rapid dissemination of recently introduced genetic material within the gene pool (14, 15). Other families of phages frequently exhibiting significantly different lifestyles such as the Mu family (70) (see chapter 30) or N15 (75) (see chapter 28), maintain a different genome organization to the lambdoid family (see chapter 27). Examination of the organization of the *Streptomyces* phages suggests that they might also be classified into families. The organization of the ϕC31 genome is known in detail and is described below. However, even before the ϕC31 sequence was available an overview of its genome organization was gleaned from the position of the *attP*/*int* locus and the locations of the deletable regions in the temperate phages (18, 19, 43, 64). In phages ϕC31, ϕA7, and TG1 the *attp*/*int* lies close to one of the *cos* ends (18, 31, 34). In all three phages the deletable regions are either adjacent to the *attP*/*int* region, (the inessential early region) or in the center of the genome, the latter being associated with a clear-plaque phenotype (the *c* locus). Some limited sequencing of restriction fragments from TG1 confirmed the similarity of genome organization between TG1 and ϕC31 (S. Sharp and M. C. M. Smith, unpublished data).

Phages R4, ϕA8, VWB, and RP3 appear to belong to at least one other family on these, albeit crude, criteria. The locations of *attP*/*int* in R4, RP3, and VWD are closer to the centers of the genomes compared with ϕC31 (36, 82, 100). In VWB the *int*/*attP* is just to the left of center and the coat protein genes are located on the right arm. A derivative of SH10 (a phage that has a similar restriction map to R4) has a deletion close to one end of the genome and under certain plating conditions has a clear-plaque phenotype (102); if this deletion indicates the position of the repressor gene, its position is very different in phage SH10 compared with that in ϕC31. Other phages cannot be classified with these criteria; ϕHAU3 also has *attP* at one end but insertions to generate phagemids have occurred at the other end of the genome in a position that in ϕC31 would coincide with the essential terminase and coat genes (107). No information is available for other phages, but on the basis of genome size alone it is highly likely that the FP43, FP22, ϕA1, and DAH phages belong to yet other families (table 39-1).

Streptomyces Temperate Phage ϕC31

The most extensively characterized phage of *Streptomyces* spp. is the temperate phage ϕC31. This phage was isolated in the late 1960s, ostensibly from a culture supernatant of *S. coelicolor* A3(2) plated on *S. lividans* 66 (63). Subsequent studies indicated that ϕC31 must have come from external contamination (18). ϕC31 has a polyhedral head, a noncontractile tail, and a baseplate from which short tail fibres protrude (95) (figure 39-1). Lomovskaya and colleagues studied adsorption, stability, and one-step growth curves, mapped the ϕC31 chromosomal attachment site (*attB*), and undertook a genetic analysis of the phage genome. Moreover, they discovered the phage growth limitation phenotype (Pgl) phenotype in *S. coelicolor*. The following sections expand on these discoveries.

Genome Organization of ϕC31

The complete genome sequence of ϕC31 has been determined (44, 61, 92). The ϕC31 genome is 41,491 bp in length and contains 63.6% G+C (figure 39-2). The arrangement of genes indicates that transcription of all genes except one is in the left to right orientation. As in other phage genomes, the early and late genes are clustered, with the late genes on the left arm and the early genes on the right arm. This was first suggested from the genetic analysis of temperature-sensitive (ts) mutants of ϕC31 later confirmed by S1 mapping (53, 65, 96). There is a further subclustering of genes according to function, for example head and tail assembly, DNA packaging, and DNA replication (figure 39-2). Comparisons of phage genomes have indicated that the

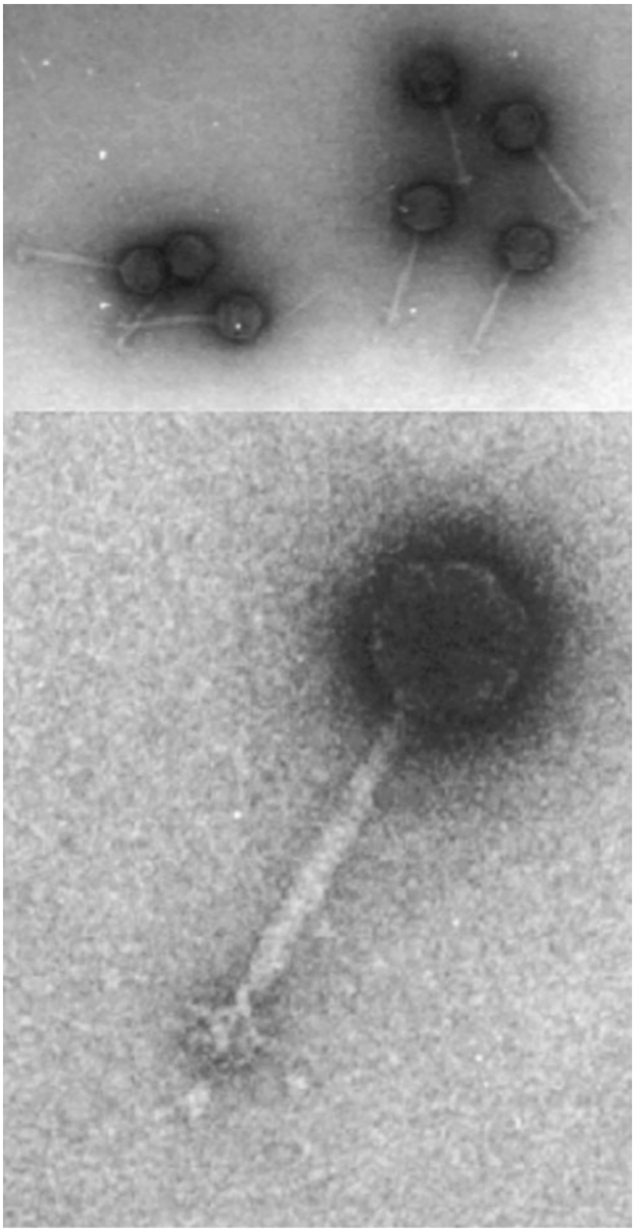


Figure 39-1 Electron micrographs of phage ϕC31. Micrograph of R. Hendrix, Pittsburgh Bacteriophage Institute.

mosaicism of sequence relatedness between phages is not random; the transition point between sequence similarity and divergence frequently lay at the boundaries of so-called functional modules (15, 54, 99). If genes for a particular function are clustered and genetic exchange occurs, recombinants are less likely to lack an essential component than if the genes are scattered.

Overall the late genes of ϕC31 have a similar gene organization to many other phage late regions (9). The major capsid and tail coat proteins were identified by N-terminal sequencing and located on the gene map of ϕC31 (92). Other structural genes encoding coat proteins or putative packaging genes were identified by homology to other phage genes in the database. The gene products in the ϕC31 head assembly module are similar to those in a group of phages that include coliphage HK97 (92). The pathway for head assembly has been studied in detail in HK97 (25, 35, 104) and it is likely that it is similar in ϕC31. The arrangement of the tail genes also resembled that of many other phages and includes the presence of the putatively frameshifted open reading frame (ORF), g42, and adjacent to this, a putative tail tape measure gene, g43 (92). The product of g49 was also of some interest due to the presence of collagen-like repeats, and this is found to be quite common in putative tail fiber proteins (89). The early regions are generally more variable between different phages than the late regions (9). In phage ϕC31, g11, which is found in the early region, encodes a putative DNA polymerase. Although considered unusual for most temperate phages, genes encoding DNA polymerases are common in the mycobacteriophages (44, 73) (which are reviewed in chapter 38). Adjacent to *g11* in ϕC31 is *g9a* encoding a putative primase/helicase similar to many other phage-encoded primases including the α protein from phage P4 (109) (phage P4 is reviewed in chapter 26). It is tempting to speculate that the mode of replication might be similar to P4, that is bidirectional replication beginning at either of two origins (6). Whilst still on the theme of DNA synthesis, ϕC31 and some mycobacteriophages appear to encode enzymes for the metabolism of nucleotides. In ϕC31, *g15* through *g28* and *int* lie within the region termed "inessential", as deletions can be generated in this region without affecting plaque formation (17, 92). This region contains two genes: *g20*, encoding a putative dCMP deaminase with close similarity to gene *36.1* from mycobacteriophage D29, and *g16*, encoding a putative alternative thymidylate synthase and similar to *g48* from both L5 and D29. ϕC31 also contains *g52* which encodes a putative nucleotide kinase, most similar to the T4-encoded homolog (phage T4 is reviewed in chapter 18). Although the idea has never been tested, all three of these genes may be advantageous to ϕC31 replication by modifying the nucleotide pools during lytic growth to maximize the nucleotide pools during lytic growth to maximize the burst size (see chapter 5 for a discussion of phage growth parameters such as burst size).

The genome of phage ϕBT1 is extremely similar to that of phage ϕC31 with most of the genes in the late arm encoding proteins between 73% and 96% identical to the ϕC31 homolog (39). The early regions contain more variation, with amino acid sequence homologies between <30% to 92% identity. Moreover, several genes are present in one phage without a corresponding homolog being present in the other. One of these is ϕC31 *g16* (discussed above), which is predicted to encode an alternative thymidylate synthase and is not present or required in phage ϕBT1. On the other hand ϕBT1, but not ϕC31, encodes a putative DNA methyltransferase, located between *g9a* and *g11*, two genes that,

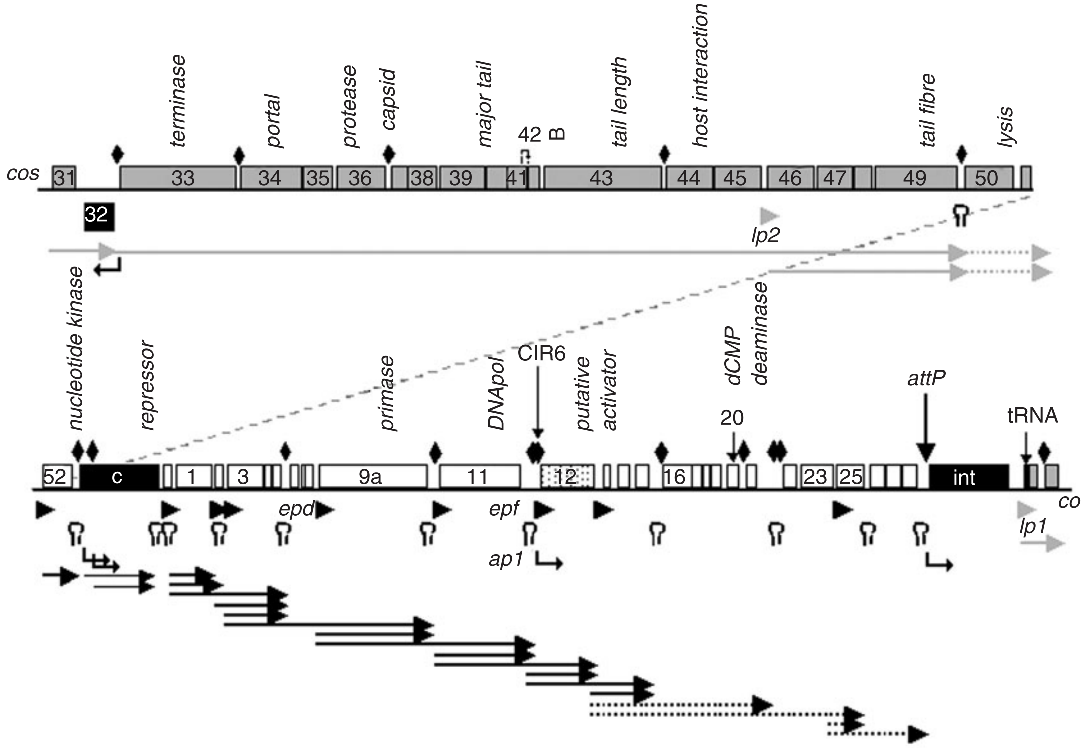


Figure 39-2 Genomic organization and transcription map of phage ϕC31. The genome is represented by a line, broken at the middle, with the left arm being shown in the upper section and the right arm in the lower section. The ends represent the cohesive (*cos*) ends of the linear genome. The genes are represented by boxes, some of which are numbered, with the late genes in gray, early genes in white, and immediate early genes (repressor, *c*; integrase, *int*; gene *32*) in black. The putative activator, gene *12*, is represented by a stippled box. The arrows below the genome represent transcripts; black are early or immediate early and gray are late. The dotted lines are putative transcripts that have not been directly observed by S1 mapping. The positions of early promoters (black arrowheads), late promoters (gray arrowheads), immediate early promoters (bent arrows) and terminators (stem-loop icons) are shown. Diamonds represent the repressor binding or conserved inverted repeat (CIR) sites.

respectively, encode homologs of DNA primase and DNA polymerase (figure 39-2; 39). This additional gene in ϕBT1, *g9.1*, could be part of an antirestriction strategy for ϕBT1 or alternatively part of a superinfection exclusion system for ϕBT1 lysogens. Downstream of *g9.1* is *g9.2*, also not present in ϕC31, a gene that contains no recognizable functional motifs. However, these two genes, *g9.1* and *g9.2*, have homologs in mycobacteriophage Corndog, that is *g7* and *g6* (73) (see chapter 38), and the fact that they are found together in both ϕBT1 and in Corndog suggests that the products of these two genes may function together.

Infection

The infection process requires adsorption of the phage particle to a cell wall receptor and then injection of DNA into the host cell. Lomovskaya and colleagues studied the relationship between adsorption and successful infection of *S. coelicolor* (65). Phage particles reversibly adsorb to spores. Only germlings of about 5 hours old yield productive infections, however, and these represent only about 20% of the particles absorbed to germlings. Recent studies have shed light on the nature of the ϕC31 receptor (27). Mutants of *S. coelicolor* strain J1929 (a stable Pgl^{-} mutant:see below) resistant to ϕC31 were isolated by UV or spontaneous mutation and shown to plate the phage at extremely low efficiency. These strains could support a phage burst if the DNA was introduced by transfection, indicating that the defect was early in the infection cycle. Several of the phage-resistant strains, however, could support growth of a ϕC31 derivative isolated originally in Lomovskaya's laboratory called ϕC31*hc*. This phage was itself isolated as being able to grow on a phage-resistant derivative of *S. lividans* and it seems likely that phage ϕC31*hc* has compensated for alteration of the host receptor. Restoration to phage sensitivity was achieved in one of the resistant *S. coelicolor* strains, DT1017, by the introduction of a gene, SCO3154, encoding a homolog to the eukaryotic dolichol phosphate-D-mannose:protein *O*-D-mannosyltranferase (27). Phage sensitivity could be restored to other phage-resistant mutants by the introduction of SCO1423, a gene encoding a homolog of dolichol phosphate-β-D-mannose synthase (26).

Together these findings strongly indicate that the receptor for wild-type ϕC31*hc* is a glycoprotein. Possibly ϕC31*hc* can use the unmodified or incompletely modified cell wall protein, which is yet to be identified, as a receptor. In support of this model, a phage, ϕDT4002, was isolated by purification of a derivative of ϕC31cΔ25 that can grow with a high efficiency of plating on *S. coelicolor* DT1017 (i.e., the same phenotype as ϕC31*hc*) and the late genes sequenced; compared with its direct parent, phage ϕC31*c*Δ25, *g44* in ϕDT4002 contained a missense mutation, implicating gp44 in interactions with the host receptor (figure 39-2).

Induction into Lytic Growth

S. coelicolor lysogens are not sensitive to UV (76). Indeed, the physiological conditions that stimulate induction of lysogens into lytic growth are not known although high levels of induction are observed during growth of young mycelia (76). Indeed, ts lysogens of *S. coelicolor* could not produce progeny phage if induced before 3 hours or after 10 hours post-induction (76). The inability of aging mycelia to produce phage particles was studied further by Suarez and colleagues and found to be due to a block in transcription (96). Thus, ϕC31 appears to have a mechanism to ensure maximal phage growth during rapid vegetative growth of the host. Interestingly, studies with phage–host interactions in soil confirm these observations (see below).

In contrast, a lytic phage, JHJ-1, that infects *Saccharopolyspora hirsuta* appears to be able to grow on aging mycelia (42). The plaques formed by JHJ-1 consequently are not self-limiting and continue to grow for several days, a phenotype reminiscent of coliphage T7 plaque growth (see chapter 20).

Transcription in the Lytic Cycle

Genetic loci required for lytic growth were characterized by the isolation of ts mutants for phage replication (65). The time at which growth was sensitive to temperature indicated that all except one isolate were mutants in loci required for late lytic growth. These loci were mapped by 2- and 3-factor crosses; late genes lay to the left of the clear-plaque locus and the single early mutation mapped to the right. S1 mapping experiments confirmed this organization using RNA prepared from induced ts lysogens containing a ϕC31 *cts1* prophage containing a mutant allele of the repressor locus, *c* (53, 96). Fine-mapping of the early region showed that discrete transcripts were maximal 10 minutes post-induction and were almost absent by 20 minutes (figure 39-2) High-resolution mapping of the 5′ endpoints showed that transcription in the early region initiates at multiple promoters, and a highly conserved sequence was identified in all these promoters (51). Late genes are expressed from a promoter, *lp1*, at the start of the late operon and transcription probably continues through to a terminator located upstream of *g50* (figure 39-2; 48, 49). A second late promoter, *lp2*, is located in the late operon.

Using transcriptional fusions the timing of transcription of the phage promoters was examined. An early promoter, *epd*, and a late promoter, *lp1*, were separately fused to *xylE* and introduced into a ϕC31*cts* lysogen (48, 51). Temperature induction synchronously induced the phage lytic cycle catechol 2,3-dioxygenase activity was maximal at 20 and 40 minutes, respectively, for the early and late promoters. Neither promoter was active in the absence of induction. The phage promoters are probably activated by a phage-encoded protein and circumstantial evidence implicates the product of *g12* as being the putative activator. *g12* is expressed from two promoters: an immediate-early promoter, *ap1*, and a phage-specific early promoter, *epf*. Thus *g12* is transcribed immediately upon entering the cell and then transcription continues once the lytic cycle is under way due to a proposed auto-activation via *epf*.

The promoter *ap1* is repressed by an operator, CIR6, which is a binding site for the phage repressor (see below) and this provides a mechanism for controlling the lysis versus lysogenic decision (105). Furthermore, a virulent mutant of ϕC31, ϕC31*vir1*, was shown to contain a DNA rearrangement, resulting in the complete separation of the −10 promoter sequence of *ap1* from the CIR6 operator. The resulting unregulated expression of *g12* could explain the virulent phenotype (105) Unfortunately, all attempts to clone *g12* in the absence of the phage repressor have failed and we propose that this is due to toxicity of the *g12* gene product.

Studies directed at host transcription during the ϕC31 lytic cycle have shown a remarkable reduction in rRNA synthesis and possibly transcription of all host genes (23, 76). This type of phenomenon is more reminiscent of lytic phages such as T4 (see chapter 18). Although the mechanism for the reduction in host rRNA synthesis is not clear, it does require protein synthesis after the start of the lytic cycle, consistent with the synthesis of a phage-encoded protein. These studies also noted a stimulation of ϕC31 transcribing activity in crude preparations of RNA polymerase prepared from induced cultures, perhaps suggesting that the activator of phage promoters (putatively gp12) is associated with RNA polymerase (23).

Lysogenic Growth

Lomovskaya and colleagues isolated clear-plaque mutants, including ts and dominant *c* mutants as well as the very rare virulent mutants, which do not grow on bacteria lysogenic for ϕC31 (65). All the clear-plaque mutants mapped to a single locus, *c*, located in the center of the phage genome (85). The repressor gene was isolated by its ability to confer ϕC31 resistance (83). Transcription mapping and analysis of the expression of the repressor locus indicated that three in-frame proteins—the 74, 54, and 42 kDa isoforms—are naturally produced from the *c* locus (84, 91). Plasmids that could only express the 54 and 42 kD a isoforms conferred full immunity to superinfecting phage. Purified 42 and 54 kDa isoforms were found to bind to a conserved inverted repeat sequence (CIR) located at 16 places in the phage genome (52, 105, 106). One of these CIRs appears to be crucial for regulation of *g12* and therefore the lytic cycle (see above). Others appear to autoregulate the repressor (52, 84). The roles of the remaining sites are not clear. The role of the 74 kDa repressor is also not understood, although it has been suggested to act as an antirepressor (90).

For most temperate phages the establishment of lysogeny requires two events:the repression of lytic genes and the site-specific recombination of the phage genome into a particular target (*attB*) in the host genome (see chapters 7 and 8). The ϕC31, ϕBT1, and R4 recombination systems are unusual compared with most known phage recombination systems as they employ integrases from the serine recombinase family and are unrelated to the phage λ integrase-like recombinases of RP3 and VWB (24, 36, 40, 61, 82, 100). Phage-encoded integration and excision is a highly regulated process to ensure that the reaction is appropriate to lytic or lysogenic growth. In phage λ, transcriptional and post-transcriptional mechanisms operate to ensure that the appropriate ratio of Xis to Int is expressed for integration and excision. Both the RP3 and VWB recombination systems also have recognizable *xis* genes located adjacent to the *int* genes (36, 100). In vitro studies on the properties of ϕC31 integrase show that the recombinase causes integrative recombination with *attP* and *attB* but not excisive recombination (see chapter 7 for a general discussion of phage integration and excision). Thus, phage ϕC31 is also thought to encode a Xis protein to direct excision by integrase, but this has not yet been identified. An interesting aspect of the control of ϕC31 *int* expression is that the *int* promoter straddles the crossover site (61). After integration the promoter is therefore separated from the *int* ORF. Moreover, the *attB* site lies within a host ORF, SCO3798, and the direction of integration precludes transcription of *int* from the promoter of this host gene.

Restriction and Antirestriction in Phage Infection

It is generally thought that restriction of foreign DNA is the host property that, in addition to adsorption sensitivity, most limits the phage host range (28). Phage-sensitive mutants have been isolated from resistant strains and this has revealed the presence of type II restriction systems (e.g., *Streptomyces albus* G encoding *Sal*I and *S. albus* P encoding *Sal*PI) and the unusual phage defence system, Pgl, in *S. coelicolor* A3(2). The high selection pressure to overcome restriction in phage genomes has sometimes resulted in removal of sequences that act as targets for restriction endonucleases (81). This has been shown to be true for several *Streptomyces* phages (see data compiled in 17). This is very much in line with the empirical findings of Cox and Baltz (28).

Now that we have the sequences of two complete *Streptomyces* phage genomes, the analysis can be more precise. Both ϕC31 and ϕBT1 sequences show a severe reduction in the expected number of sites for several enzymes produced by *Streptomyces* species whilst others, such as *Sal*I, *Sap*I, *Sfu*I, *Ssb*I, and *Ssp*5230I, have considerably higher than expected numbers (table 39-2). Possibly the high frequency of sites for these enzymes reflects the presence of antirestriction mechanisms, and it is of interest that one phage, FP43, does appear to display antirestriction (69). Furthermore, the antimodification system displayed by phages R4 and ϕC31 to *Sal*I may also be part of the avoidance of restriction (see below). Of particular interest is the number of *Sac*AI sites in ϕC31 versus ϕBT1; the former appears to have removed *Sac*AI sites and the latter appears to have increased their number, suggesting the presence of a mechanism conferring immunity against *Sac*AI restriction in ϕBT1. The sequence of ϕBT1 reveals the presence of a putative methyltransferase, absent in ϕC31 and which may be part of an antirestriction strategy (39). Another point that was clear from the analysis by Chater (17) is that several of the restriction sites that appear to have been avoided by *Streptomyces* phages are for enzymes that are common in many *Streptomyces* species, perhaps the best examples being CTCGAG and CCGCGG (table 39-2).

Streptomyces phage FP22 has an extremely effective mechanism for evading restriction. This Phage appears to contain none of the sites for characterized *Streptomyces* type II restriction systems, even *Mbo*I recognition sites (41). Possibly this reflects a general modification of phage DNA.

The "Antimodifying" *g* Locus in R4 and ϕC31

Phages R4 and ϕC31 cannot form plaques on *S. albus* G but can on a R^-M^+ mutant of *S. albus* (17, 20). Surprisingly, however, the phages derived from the R^-M^+ mutant are still restricted on *S. albus* G. Mutants can be isolated, called *g* mutants, that will plaque efficiently on *S. albus* G. DNA isolated from g^+ R4 or ϕC31 grown on the R^-M^+ strain are still restricted with *Sal*I but the g^- mutants are modified and are resistant to cleavage. The phenotype of the wild-type phage therefore appears to be antimodifying, such that the phage can grow in an R^-M^+ host but does not become modified. Now that it is clear that many *Streptomyces* host strains contain methylation-specific restriction systems (MSR), it appears to make sense that the phage resists modification in case its next host carries an MSR system (66). Indeed, it has been shown that R4 g^- but not R4 g^+ was

Table 39-2 Frequencies of Recognition Sites in ϕC31 and ϕBT1 for *Streptomyces* and *Saccharopolyspora* Restriction

Enzyme[a]	Source	Isoschizomers from other *Streptomyces* spp.	ϕC31	ϕBT1	Recognition site	Expected
*Sac*AI	*S. achromogenes* ATCC 21353	6	2	206	GCCGGC	43
*Sac*I	*S. achromogenes* ATCC 12767	1	0	0	GAGCTC	14
*Sac*II	*S. achromogenes* ATCC 12767	15	2	0	CCGCGG	43
*Sac*NI	*S. achromogenes* N-J-H		159	35	GRGCYC	106
*Sal*AII	*S. albus* ATCC 21725	1	128	389	GATC	139
*Sal*DI	*S. albus* ATCC 21132		28	11	TCGCGA	14
*Sal*GI	*S. albus* G ATCC 29789		48	28	GTCGAC	14
*Sal*PI	*S. albus* P CMI52766	5	0	0	CTGCAG	14
*San*DI	*Streptomyces* sp.	1	8	11	GGGWCCC	16
*Sap*I	*Saccharopolyspora* sp. NEB 597		38	42	GCTCTTC	2.6
*Sau*HI	*S. aureofaciens* 13		0	0	CCTNAGG	14
*Sbf*I	*Streptomyces* sp. Bf-61	2	0	0	CCTGCAGG	1.4
*Sca*I	*S. caespitosus*		0	0	AGTACT	4.6
*Sco*I	*S. coelicolor* ATCC 10147		0	0	GAGCTC	14
*Sex*AI	*S. exfoliates*		1	1	ACCWGGT	5.1
*Sfi*I	*S. fimbriatus* ATCC 15051		0	1	GGCCNNNNNGGCC	4.3
*Sfr*274I	*S. fradiae* 274	22	0	0	CTCGAG	14
*Sfu*I	*S. fulvissimus*	10	15	21	TTCGAA	4.6
*Sgf*I	*S. griseoruber*		0	0	GCGATCGC	1.4
*Sgh*18351	*S. ghanaensis* 1835		94	84	GGWCC	154
*Sgr*201	*S. griseus* K20	1	108	115	CCWGG	154
*Sgr*AI	*S. griseus*		2	3	CRCCGGYG	10.7
*Slb*I	*S. albidoflavus* Viikki 329		2	7	GGTCTC	14
*Sno*I	*S. novocastria*		9	12	GTGCAC	14
*Sol*3335I	*S. olivaceus* IMRU3335		0	0	CAGCTG	14
*Sph*I	*S. phaeochromogenes* NRRL B-3559	1	8	10	GCATGC	14
*Spv*I	*S. parvus* NRRL B-1255		0	0	GGATCC	14
*Srf*I	*Streptomyces* sp.		0	0	GCCCGGGC	4.3
*Ssb*I	*S. scabies*		17	19	AAGCTT	4.6
*Sse*232I	*Streptomyces* sp. RH232		0	3	CGCCGGCG	4.3
*Sse*86471	*Streptomyces* sp. 8647		1	2	AGGWCCT	5.1
*Sse*AI	*Streptomyces* sp.		64	58	GGCGCC	43
*Ssp*2I	*Streptomyces* sp. RFL2		259	239	CCSGG	270
*Ssp*5230I	*Streptomyces* sp. 5230		53	52	GACGTC	14
*Ssp*BI	*Streptomyces* sp.	1	7	4	TGTACA	4.6
*Sst*IV	*S. stanford*		19	14	TGATCA	4.6
*Stu*I	*S. tubercidicus*	4	0	0	AGGCCT	14

[a] Information on restriction endonucleases from *Streptomyces* sp. is from http://rebase.neb.com/rebase/

restricted by *Streptomyces parvulus* after propagation on *S. fradiae* NRRL F1144 (17). It is now known that *S. parvulus* contains an MSR system (66) so if *S. fradiae* is a modifying host then only the g^+ phages (antimodified) should avoid restriction. The *g* locus has been mapped by deletion analysis and by recombination and shown to be located toward the left of the inessential region (64, 88).

ϕC31 grown in *S. albus* R^-M^+ is apparently restricted approximately 100-fold on *S. coelicolor* A3(2) whereas ϕC31 containing the Moscow deletion (ΔM) was not restricted (17). Wild-type ϕC31 therefore appears to provide a substrate for restriction that is not present in the ΔM derivatives. ΔM removes all or part of genes 19–23 in the ϕC31 inessential region (C. Finnis and M. C. M. Smith, unpublished data). Other phages that appear to be able to avoid restriction systems include ϕA4 and ϕA2. The former avoids the Sac enzymes and the latter the *Sal*P1 enzyme (32).

The Phage Growth Limitation (Pgl) Phenotype in *S. coelicolor* A3(2)

S. coelicolor A3(2) is naturally resistant to ϕC31 infection (65). Whilst this was thought initially to be due to the presence of a resident ϕC31 prophage that was conferring immunity, this was refuted when Southern analysis was performed on *S. coelicolor* genomic DNA with ϕC31 DNA as a probe; no DNA homologous to ϕC31 DNA was detected (18). The resistance phenotype was later found to be due to the novel phage defence system conferred by Pgl (21). Briefly, ϕC31 produced from a strain, such as *S. lividans*, that does not naturally possess the Pgl system (or a Pgl^- mutant of *S. coelicolor*) is able to infect wild-type Pgl^+ strains produced a normal burst of progeny phage. However, when these progeny phage are used to infect fresh Pgl^+ bacteria, the number of second-generation phages produced is greatly attenuated. In the model proposed by Chinenova et al. (21), progeny phage are modified after growth in a Pgl^+ host and this activates a mechanism to inhibit further phage replication. The modification does not inactivate phages released from Pgl^+ hosts, as they can efficiently infect a Pgl^- host to give successive rounds of normal phage bursts. One can speculate that such a strategy is particularly advantageous to a *Streptomyces* colony; the sacrifice of one cell to phage infection protects the rest of the clone but additionally amplifies the phage so that competing species in the same niche that do not encode Pgl are more likely to be infected and thereby eliminated.

It is now known that four genes are required for the Pgl phenotype, *pglWXYZ* (4, 62, 97). The *pglWX* and *pglYZ* loci are only separated by about 6 kbp but the genes between them apparently are not required. A plasmid containing *pglWX* and *pglYZ* can confer the Pgl phenotype on *S. lividans*, a close relative of *S. coelicolor* that does not naturally contain any homologs to the *pgl* genes. PglX is predicted to be a DNA adenine methyltransferase and, in support of the Chinenova model, suggests a role for DNA methylation. There are no recognized motifs in PglZ but this protein may be modified, possibly by PglW, which is predicted to be a signaling protein containing Ser/Thr protein kinase domains and a HTH DNA binding motif (47, 97). How the proteins confer Pgl is unknown at the moment, as is the target in ϕC31 on which Pgl is presumed to act. No ϕC31 derivatives have been isolated that are resistant to Pgl, suggesting that the target is either in multiple copies in the phage genome or is essential (62). As only phages in the same immunity grouping as ϕC31 were sensitive to Pgl, it was suggested that Pgl targeted the repressor-operator system, possibly the highly conserved operators or CIR sites (97). It has now been discovered that a heteroimmune phage, a derivative of ϕHAU3, is sensitive to Pgl and this is currently being investigated further (J. Leafe and M. C. M. Smith, unpublished data).

Another interesting feature of the Pgl system is a high frequency (approximately 10^{-3} to 10^{-4} per spore) of phase variation from $\mathrm{Pgl}^+ \rightarrow \mathrm{Pgl}^-$ and vice versa (62). Expansion and contraction of a G-tract within the *pglX* gene have been shown recently to account for Pgl phase variation in several strains of *S. coelicolor* (98). It was also proposed that the phase variation in the putative methyltransferase gene might be part of a mechanism that switches the target of Pgl and therefore the phage specificity. *S. coelicolor* encodes a paralog of *pglX*, called *pglS*, the product of which might be able to interact with the remaining Pgl proteins if *pglX* is not expressed and direct the Pgl phenotype to an alternative group of phage (98).

How prevalent is Pgl? Recently homologs of *pglWXYZ* have been deposited in the databases from the genome sequencing of *Thermobifida fusca*, a mycelial thermophilic actinomycete (http://www.ncbi.nlm.nih.gov/genomes/). More distant relatives of the Pgl proteins can be detected using PSI-BLAST in the cyanobacterium *Nostoc punciforme*, the archaeon *Methanosarcina*, and in *Providencia rettgeri*, *Vibrio cholera*, and *Salmonella typhimurium* LT2. As in *S. coelicolor* and *T. fusca*, the genes encoding the Pgl-like proteins in these diverse organisms are clustered and in *P. rettgeri* and *V. cholera* they are located on mobile genetic elements. However, it is unknown whether any of these *pgl*-like clusters confers phage resistance.

Other instances of possible Pgl-like phenotypes have been previously observed. *Saccharopolyspora erythraea* NRRL 2359 appears to be sensitive to "lysis from without" by certain phage, that is SE-3 and ϕC69, but these phages cannot form plaques (55, 93). If a high moi of ϕC31 is dropped onto a lawn of *S. coelicolor* M145 (Pgl^+) a zone of reduced bacterial growth is observed but the phage is incapable of forming plaques.

Host–Phage Interactions in Soil

Most of our knowledge of phage–host interactions comes from studies in a very unnatural environment: agar plates on rich media (discussed generally for phages in chapter 5). How do these studies relate to the natural environment — in the case of the streptomycetes, in soil? Important observations on the persistence of phage, survival of lysogens, the infection cycle itself have been made by Prof. Wellington and coworkers (13, 29, 46, 67). As in the agar-plate experiments, rapid increases in phage counts were observed under conditions that would favor germination and mycelial outgrowth (29, 67). Thus phage multiplication was observed in a soil environment on inoculation of fresh soil with spores of lysogens — conditions which would support rapid germination of the host and concomitant phage growth. The rise in phage numbers was followed by a fall, corresponding to the resporulation of the host. Smaller bursts were observed after addition of fresh soil or mixing. Equilibrium was soon established where the numbers of lysogens and free phage remained constant. Conversion of endemic bacteria to lysogens on inoculation of unsterile soil with spores of a lysogen occurred frequently (67), although this could be to the detriment of the inoculated lysogen (46). An unusual observation concerning the phage–*Streptomyces* interactions in soil is that the host numbers do not vary in response to the numbers of phage (13). This is at odds with observations from the marine environment (see chapter 32 and 33 as well as chapter 5). The authors attribute the stability of the host to: (i) susceptibility of newly germinated spores rather than older mycelium, (ii) substrate mycelium in mature colonies that adsorb approximately 98% of total phages thus protecting young mycelia, (iii) a large burst size in soil, and (iv) no measurable impact on growth of the host by the phage, possibly due to the spatial heterogeneity in the soil environment (13).

Outlook

As in other fields in biology, genomics will fundamentally enhance our understanding of *Streptomyces* phages. Some of the greatest opportunities for research consequently lie ahead. One important question that is currently being addressed concerns the total diversity of phage genomes. Recently snapshots of the phage gene pool from different environments have been taken (7, 8). Virus-sized particles were purified directly from the environment and fragments of their DNA sequenced without imposing a bottleneck through amplification in specific bacterial hosts. The results indicate tremendous diversity, with up to 65% of the sequences being unique. Another study comparing the genomes of 14 randomly isolated mycobacteriophages (see chapter 38) upholds this view of unprecedented diversity (73). There is no reason not to believe that the *Streptomyces* phages will be equally diverse. Genomics combined with ecological and laboratory-based approaches will continue to address the uniqueness of *Streptomyces* phage biology. What are the adaptations that *Streptomyces* phages have undergone that enable their growth and survival in their unusual hosts? How have phage evolved to an organism that has a mycelial growth habit and are there any phage-encoded genes that enhance or modify antibiotic synthesis? Conversely, how has the mycelial host evolved in response to phage infection? Finally, what role do phages have in the transfer of genetic information in the soil environment and, in the case of *Streptomyces*, in the evolution of antibiotic biosynthesis pathways?

Acknowledgments

I am grateful to Prof. Keith Chater for his endless enthusiasm and useful suggestions for all aspects of our work in this field and for his comments on the manuscript. Research in M. C. M. S's laboratory is funded by the BBSRC.

References

1. Ackermann, H. W. 2001. Frequency of morphological phage descriptions in the year 2000. Brief review. Arch. Virol. 146:843–857.
2. Anne, J., P. Fiten, L. Van Mellaert, B. Joris, G. Opdenakker, and H. Eyssen. 1995. Analysis of the open reading frames of the main capsid proteins of actinophage VWB. Arch. Virol. 140:1033–1047.
3. Anne, J., P. Verheyen, G. Volckaert, and H. Eyssen. 1985. A restriction endonuclease map of *Streptomycis* phage VWB. Mol. Gen. Genet. 200:506–507.
4. Bedford, D. J., C. Laity, and M. J. Buttner. 1995. Two genes involved in the phase-variable ϕC31 resistance mechanism of *Streptomyces coelicolor* A3(2). J. Bacteriol. 177:4681–4689.
5. Bentley, S. D., K. F. Chater, A. M. Cerdeno-Tarraga, G. L. Challis, N. R Thomson, K. D. James, D. E. Harris, M. A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C. W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C. H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O'Neil, E. Rabbinowitsch, M. A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D. Saunders, S. Sharp, R Squares, S. Squares, K. Taylor, T. Warren, A. Wietzorrek, J. Woodward, B. G. Barrell, J. Parkhill, and D. A. Hopwood. 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature 417:141–147.
6. Bertani, L. E., and E. W. Six. 1988. The P2-like phages and their parasite, P4, pp. 73–143. *In* R. Calendar (ed.) The Bacteriophages, vol. 2. Plenum Press, New York.
7. Breitbart, M., I. Hewson, B. Felts, J. M. Mahaffy, J. Nulton, P. Salmon, and F. Rohwer. 2003. Metagenomic analyses of an uncultured viral community from human feces. J. Bacteriol. 185:6220–6223.
8. Breitbart, M., P. Salamon, B. Andresen, J. M. Mahaffy, A. M. Segall, D. Mead, F. Azam, and F. Rohwer. 2002. Genomic analysis of uncultured marine viral communities. Proc. Natl. Acad. Sci. USA 99:14250–14255.
9. Brussow, H., and F. Desiere. 2001. Comparative phage genomics and the evolution of Siphoviridae: insights from dairy phages. Mol. Microbiol. 39:213–222.
10. Brussow, H., and R. W. Hendrix 2002. Phage genomics: small is beautiful. Cell 108:13–16.
11. Brzezinski, R., E. Surmacz, M. Kutner, and A. Piekarowicz. 1986. Restriction mapping and close relationship of the DNA of *Streptomyces erythraeus* phages 121 and SE-5. J. Gen. Microbiol. 132:2937–2943.
12. Burke, J., D. Schneider, and J. Westpheling. 2001. Generalized transduction in *Streptomyces coelicolor*. Proc. Natl. Acad. Sci. USA 98:6289–6294.
13. Burroughs, N. J., P. Marsh, and E. M. Wellington. 2000. Mathematical analysis of growth and interaction dynamics of streptomycetes and a bacteriophage in soil. Appl. Environ. Microbiol. 66:3868–3877.
14. Campbell, A. 1988. Phage evolution and speciation, pp. 1–14. *In* R. Calendar (ed.) The Bacteriophages, vol. 1. Plenum Press, New York.
15. Casjens, S., G. F. Hatfull, and R W. Hendrix. 1992. Evolution of dsDNA tailed-bacteriophage genomes. Semin. Virol. 3:383–397.
16. Chater, K. F. 2001. Regulation of sporulation in *Streptomyces coelicolor* A3(2): a checkpoint multiplex? Curr. Opin. Microbiol. 4:667–673.
17. Chater, K. F. 1986. *Streptomyces* phages and their applications to *Streptomyces* genetics, pp. 119–157. *In*

S. W. Queener and L. E. Day (eds.) The Bacteria: A Treatise on Structure and Function, vol. 9: Antibiotic Producing *Streptomyces*. Academic Press, Orlando, Fl.

18. Chater, K. F., C. J. Bruton, A. A. King, and J. E. Suarez. 1982. The expression of *Streptomyces* and *Escherichia coli* drug-resistance determinants cloned into the *Streptomyces* phage ϕC31. Gene 19:21–32.
19. Chater, K. F., C. J. Bruton, and J. E. Suarez. 1981. Restriction mapping of the DNA of the *Streptomyces temperate* phage ϕC31 and its derivatives. Gene 14:183–194.
20. Chater, K. F., and A. T. Carter. 1979. A new, wide host-range, temperate bacteriophage (R4) of *Streptomyces* and its interaction with some restriction-modification systems. J. Gen. Microbiol. 115:431–442.
21. Chinenova, T. A., N. M. Mkrtumian, and N. D. Lomovskaia. 1982. [Genetic characteristics of a new phage resistance trait in *Streptomyces coelicolor* A3(2)]. Genetika 18:1945–1952.
22. Chung, S. T. 1982. Isolation and characterization of *Streptomyces fradiae* plasmids which are prophage of the actinophage ϕSF1. Gene 17:239–246.
23. Clayton, T. M., and M. J. Bibb. 1990. Induction of a ϕC31 prophage inhibits rRNA transcription in *Streptomyces coelicolor* A3(2). Mol. Microbiol. 4:2179–2185.
24. Combes, P., R. Till, S. Bee, and M. C. M. Smith. 2002. The *Streptomyces* genome contains multiple pseudo-*attB* sites for the ϕC31- encoded site-specific recombination system. J. Bacteriol. 184:5746–5752.
25. Conway, J. F., W. R Wikoff, N. Cheng, R. L. Duda, R. W. Hendrix, J. E. Johnson, and A. C. Steven. 2001. Virus maturation involving large subunit rotations and local refolding. Science 292:744–748.
26. Cowlishaw, D. A., and M. C. M. Smith. 2002. A gene encoding a homologue of dolichol phosphate-β-D-mannose synthase is required for infection of *Streptomyces coelicolor* A3(2) by phage ϕC31. J. Bacteriol. 184:6081–6083.
27. Cowlishaw, D. A., and M. C. M. Smith. 2001. Glycosylation of a *Streptomyces coelicolor* A3(2) cell envelope protein is required for infection by bacteriophage ϕC31. Mol. Microbiol. 41:601–610.
28. Cox, K. L., and R. H. Baltz. 1984. Restriction of bacteriophage plaque formation in *Streptomyces* spp. J. Bacteriol. 159:499–504.
29. Cresswell, N., P. R. Herron, V. A. Saunders, and E. M. Wellington. 1992. The fate of introduced streptomycetes, plasmid and phage populations in a dynamic soil system. J. Gen. Microbiol. 138:659–666.
30. Desiere, F., C. Mahanivong, A. J. Hillier, P. S. Chandry, B. E. Davidson, and H. Brussow. 2001. Comparative genomics of lactococcal phages: insight from the complete genome sequence of *Lactococcus lactis* phage BK5-T. Virology 283:240–252.
31. Diaz, L. A., C. Hardisson, and M. R Rodicio. 1991. Characterization of the temperate actinophage ϕA7 DNA and its deletion derivatives. J. Gen. Microbiol. 137:293–298.
32. Diaz, L. A., C. Hardisson, and M. R Rodicio. 1989. Isolation and characterization of actinophages infecting *Streptomyces* species and their interaction with host restriction-modification systems. J. Gen. Microbiol. 135:1847–1856.
33. Foor, F., and N. Morin. 1990. Construction of a shuttle vector consisting of the *Escherichia coli* plasmid pACYC177 inserted into the *Streptomyces cattleya* phage TG1. Gene 94:109–113.
34. Foor, F., G. P. Roberts, N. Morin, L. Snyder, M. Hwang, P. H. Gibbons, M. J. Paradiso, R. L. Stotish, C. L. Ruby, B. Wolanski, et al. 1985. Isolation and characterization of the *Streptomyces cattleya* temperate phage TG1. Gene 39:11–16.
35. Ford, M. E., G. J. Sarkis, A. E. Belanger, R. W. Hendrix, and G. F. Hatfull. 1998. Genome structure of mycobacteriophage D29: implications for phage evolution. J. Mol. Biol. 279:143–164.
36. Gabriel, K., H. Schmid, U. Schmidt, and H. Rausch. 1995. The actinophage RP3 DNA integrates site-specifically into the putative tRNA(Arg)(AGG) gene of *Streptomyces rimosus*. Nucleic Acids Res. 23:58–63.
37. Gaudreau, L. R., and C. V. Dery. 1993. A cloned replicon of *Saccharopolyspora* phages JHJ-1 and JHJ-3 is stably maintained as a plasmid in various actinomycetes. Gene 126:141–146.
38. Gaudreau, L. R., J. M. Lavoie, and C. V. Dery. 1991. Biological characterization of induced phages from *Saccharopolyspora hirsuta* 367 and comparison with phage JHJ-1. J. Gen. Microbiol. 137:2347–2352.
39. Gregory, M. A. 2000. Characterisation and evolution of homoimmune *Streptomyces* phages. PhD thesis, University of Nottingham, Nottingham, UK.
40. Gregory, M. A., R Till, and M. C. M. Smith. 2003. Integration site for *Streptomyces* phage ϕBT1 and the development of novel site-specific integrating vectors. J. Bacteriol. 185:5320–5323.
41. Hahn, D. R., M. A. McHenney, and R. H. Baltz. 1990. Characterization of FP22, a large streptomycete bacteriophage with DNA insensitive to cleavage by many restriction enzymes. J. Gen. Microbiol. 136:2395–2404.
42. Haket, J., Jr., D. Desmarais, K. Mehindate, and C. V. Dery. 1990. *Saccharopolyspora hirsuta* strain 367 releases JHJ-1, a bacteriophage capable of propagation on old mycelium. J. Gen. Microbiol. 136:573–579.
43. Harris, J. E., K. F. Chater, C. J. Bruton, and J. M. Piret. 1983. The restriction mapping of *c* gene deletions in *Streptomyces* bacteriophage ϕC31 and their use in cloning vector development. Gene 22:167–174.
44. Hartley, N. M., G. O. Murphy, C. J. Bruton, and K. F. Chater. 1994. Sequence of the essential early region of ϕC31, a temperate phage of *Streptomyces* spp. with unusual features in its lytic development. Gene 147:29–40.
45. Hendrix, R. W., M. C. M. Smith, R. N. Burns, M. E. Ford, and G. F. Hatfull. 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. Proc. Natl. Acad. Sci. USA 96:2192–2197.
46. Herron, P. R., and E. M. Wellington. 1994. Population dynamics of phage–host interactions and phage conversion of streptomycetes in soil. FEMS Microbiol. Ecol. 14:25–32.

47. Hesketh, A. R., G. Chandra, A. D. Shaw, J. J. Rowland, D. B. Kell, M. J. Bibb, and K. F. Chater. 2002. Primary and secondary metabolism, and post-translational protein modifications, as portrayed by proteomie analysis of *Streptomyces coelicolor*. Mol. Microbiol. 46:917–932.
48. Howe, C. W., and M. C. M. Smith. 1996. Characterization of a late promoter from the *Streptomyces* temperate phage ϕC31. J. Bacteriol. 178:2127–2130.
49. Howe, C. W., and M. C. M. Smith. 1996. Gene expression in the *cos* region of the *Streptomyces* temperate actinophage ϕC31. Microbiology 142:1357–1367.
50. Hutchinson, C. R. 1998. Combinatorial biosynthesis for new drug discovery. Curr. Opin. Microbiol. 1:319–329.
51. Ingham, C. J., H. J. Crombie, C. J. Bruton, K. F. Chater, N. M. Hartley, G. J. P. Murphy, and M. C. M. Smith. 1993. Multiple novel promoters from the early region in the *Streptomyces* temperate phage ϕC31 are activated during lytic development. Mol. Microbiol. 9:1267–1274.
52. Ingham, C. J., C. E. Owen, S. E. Wilson, I. S. Hunter, and M. C. M. Smith. 1994. An operator associated with autocregulation of the repressor gene in actinophage ϕC31 is found in highly conserved regions in the phage genome. Nucleic Acids Res. 22:821–827.
53. Ingham, C. J., and M. C. M. Smith. 1992. Transcription map of the early region of the *Streptomyces* bacteriophage ϕC31. Gene 122:77–84.
54. Juhala, R. J., M. E. Ford, R. L. Duda, A. Youlton, G. F. Hatfull, and R. W. Hendrix. 2000. Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. J. Mol. Biol. 299:27–51.
55. Katz, L., S. J. Chiang, J. S. Tuan, and L. B. Zablen. 1988. Characterization of bacteriophage ϕC69 of *Saccharopolyspora erythraea* and demonstration of heterologous actinophage propagation by transfection of *Streptomyces* and *Saccharopolyspora*. J. Gen. Microbiol. 134:1765–1771.
56. Khosla, C., and R. J. Zawada 1996. Generation of polyketide libraries via combinatorial biosynthesis. Trends Biotechnol. 14:335–341.
57. Kieser, T., M. J. Bibb, M. J. Buttner, K. F. Chater, and D. A. Hopwood. 2000. Practical *Streptomyces* Genetics. The John Innes Foundation, Norwich.
58. Kinner, E., D. Pocta, S. Stroer, and H. Schmieger. 1994. Sequence analysis of cohesive ends of the actinophage RP3 genome and construction of a transducible shuttle vector. FEMS Microbiol. Letts. 118:283–290.
59. Klaus, S., H. Krugel, F. Suss, M. Neigefind, I. Zimmermann, and U. Taubeneck. 1981. Properties of the temperate actinophage SH10. J. Gen. Microbiol. 123:269–279.
60. Klaus, S., H. Triebel, M. Hartmann, A. Walter, F. Walter, P. Zopel, and Bar. 1979. Molecular characterization of the genomes of actinophages SH3, SH10, SH11, and SH12 infecting *Streptomyces hygroscopicus*. Mol. Gen. Genet. 172:319–327.
61. Kuhstoss, S., and R. N. Rao. 1991. Analysis of the integration function of the streptomycete bacteriophage ϕC31 J. Mol. Biol. 222:897–908.
62. Laity, C., K. F. Chater, C. G. Lewis, and M. J. Buttner. 1993. Genetic analysis of the ϕC31-specific phage growth limitation (Pgl) system of *Streptomyces coelicolor* A3(2). Mol. Microbiol. 7:329–336.
63. Lomovskaya, N., N. M. Mkrtumian, and N. L. Gostimskaya. 1970. Isolation and characterisation of *Streptomyces coelicolor* actinophage. Genetika 6:135–137.
64. Lomovskaya, N., I. Sladkova, O. Klochkova, A. V. Orekhov, T. A. Chinenova, and N. M. Mkrtumian. 1983. Genetic approaches to the development of phage cloning vectors in *Streptomyces*, pp. 66–70. *In* Genetics of Industrial Microorganisms. Kodansha, Tokyo.
65. Lomovskaya, N. D., K. F. Chater, and N. M. Mkrtumian. 1980. Genetics and molecular biology of *Streptomyces* bacteriophages. Microbiol. Rev. 44:206–229.
66. MacNeil, D. J. 1988. Characterization of a unique methyl-specific restriction system in *Streptomyces avermitilis*. J. Bacteriol. 170:5607–5612.
67. Marsh, P., I. K. Toth, M. B. Meijer, M. B. Schilhabel, and E. M. Wellington. 1993. Survival of the temperate actinophage ϕC31 and *Streptomyces lividans* in soil and the effects of competition and selection on lysogens. FEMS Microbiol. Ecol. 13:13–22.
68. Matsuura, M., T. Noguchi, D. Yamaguchi, T. Aida, M. Asayama, H. Takahashi, and M. Shirai. 1996. The *sre* gene (ORF469) encodes a site-specific recombinase responsible for integration of the R4 phage genome. J. Bacteriol. 178:3374–3376.
69. McHenney, M. A., and R. H. Baltz. 1988. Transduction of plasmid DNA in *Streptomyces* spp. and related genera by bacteriophage FP43. J. Bacteriol. 170:2276–2282.
70. Morgan, G. J., G. F. Hafull, S. Casjens, and R. W. Hendrix. 2002. Bacteriophage Mu genome sequence: analysis and comparisons with Mu-like prophages from *Haemophilus*, *Neisseria* and *Deinococcus*. J. Mol. Biol. 317:337–359.
71. Morino, T., and H. Takahashi. 1996. Transduction of a cosmid with the R4 phage *cos* sequence by heterogeneous actinophages, SPA10 and SPA38. Biosci. Biotech. Biochem. 60:2076–2077.
72. Ogata, S., H. Suenaga, and S. Hayashida. 1985. A temperate phage of *Streptomyces azureus*. Appl. Environ. Microbiol. 49:201–204.
73. PeduIla, M. L., M. E. Ford, J. M. Houtz, T. Karthikeyan, C. Wadsworth, J. A. Lewis, D. Jacobs-Sera, J. Falbo, J. Gross, N. R. Pannunzio, W. Brucker, V. Kumar, J. Kandasamy, L. Keenan, S. Bardarov, J. Kriakov, J. G. Lawrence, W. R. Jacobs, Jr., R. W. Hendrix, and G. F. Hatfull. 2003. Origins of highly mosaic mycobacteriophage genomes. Cell 113:171–182.
74. Rausch, H., M. Vesligaj, D. Pocta, G. Biukovic, J. Pigac, J. Cullum, H. Schmieger, and D. Hranueli. 1993. The temperate phages RP2 and RP3 of *Streptomyces rimosus*. J. Gen. Microbiol. 139:2517–2524.
75. Ravin, V., N. Ravin, S. Casjens, M. E. Ford, G. F. Hatfull, and R. W. Hendrix. 2000. Genomic sequence and analysis of the atypical temperate bacteriophage N15. J. Mol. Biol. 299:53–73.
76. Rodriguez, A., J. L. Caso, C. Hardisson, and J. E. Suarez. 1986. Characteristics of the developmental cycle of actinophage ϕC31. J. Gen. Microbiol. 132:1695–1701.

77. Rodriguez, E., and R. McDaniel. 2001. Combinatorial biosynthesis of antimicrobials and other natural products Curr. Opin. Microbiol. 4:526–534.
78. Schneider, J. 1989. Distribution of modules among the central regions of the genomes of several actinophages of *Faenia* and *Saccharopolyspora*. J. Gen. Microbiol. 135:1671–1678.
79. Schneider, J., I. Aguilera Garcia, and H. J. Kutzner. 1987. Characterization of a family of temperate actinophages of *Faenia rectivirgula*. J. Gen. Microbiol. 133:2263–2268.
80. Schneider, J., F. Korn-Wendisch, and H. J. Kutzner. 1990. Phi SC623, a temperate actinophage of *Streptomyces coelicolor* Muller, and its relatives ϕSC347 and ϕSC681. J. Gen. Microbiol. 136:767–772.
81. Sharp, P. M. 1986. Molecular evolution of bacteriophages: evidence of selection against the recognition sites of host restriction enzymes. Mol. Biol. Evol. 3:75–83.
82. Shirai, M., H. Nara, A. Sato, T. Aida, and H. Takahashi. 1991. Site-specific integration of the actinophage R4 genome into the chromosome of *Streptomyces parvulus* upon lysogenization. J. Bacteriol. 173:4237–4239.
83. Sinclair, R. B., and M. J. Bibb. 1988. The repressor gene (*c*) of the *Streptomyces* temperate phage ϕC31: nucleotide sequence, analysis and functional cloning. Mol. Gen. Genet. 213:269–277.
84. Sinclair, R. B., and M. J. Bibb. 1989. Transcriptional analysis of the repressor gene of the temperate *Streptomyces* phage ϕC31. Gene 85:275–282.
85. Sladkova, I., L. G. Vasil'chenko, N. D. Lomovskaya, and N. M. Mkrtumian. 1980. Physical mapping of the DNA of actinophages of *Streptomyces coelicolor* A3(2). Localisation of the *c*-region of actinophage ϕC31. Mol. Biol. 14:910–915.
86. Sladkova, I. A., T. A. Chinenova, L. B. Pel'ts, B. A. Rebentish, and N. D. Lomovskaia. 1981. [Physical mapping of *Streptomyces coelicolor* A3(2). V. Structural modifications of actinophage ϕC43 DNA molecules]. Mol. Biol. (Mosk) 15:1051–1058.
87. Sladkova, I. A., O. A. Klochkova, T. A. Chinenova, and N. D. Lomovskaia. 1984. [Physical mapping of *Streptomyces coelicolor* A3(2) actinophages. VII. Formation of deletions in the region of ϕC43 insertion sequence]. Mol. Biol. (Mosk) 18:497–503.
88. Sladkova, I. A., O. A. Klochkova, and N. D. Lomovskaia. 1982. [Physical mapping of actinophage *Streptomyces coelicolor* A3(2). VI. The use of deletion mutants of actinophage ϕC31 for construction of phage vectors]. Mol. Biol. (Mosk) 16:739–744.
89. Smith, M. C., N. Burns, J. R Sayers, J. A. Sorrell, S. R. Casjens, and R. W. Hendrix. 1998. Bacteriophage collagen. Science 279:1834.
90. Smith, M. C. M., C. J. Ingham, C. E. Owen, and N. T. Wood. 1992. Gene expression in the *Streptomyces* temperate phage ϕC31. Gene 115:43–48.
91. Smith, M. C. M., and C. E. Owen. 1991. Three in-frame N-terminally different proteins are produced from the repressor locus of the *Streptomyces* bacteriophage ϕC31. Mol. Microbiol. 5:2833–2844.
92. Smith, M. C. M., R. N. Burns, S. E. Wilson, and M. A. Gregory. 1999. The complete genome sequence of the *Streptomyces* temperate phage ϕC31: evolutionary relationships to other viruses. Nucleic Acids Res. 27:2145–2155.
93. Smorawinska, M., F. Denis, C. V. Dery, P. Magny, and R. Brzezinski. 1988. Characterization of SE-3, a virulent bacteriophage of *Saccharopolyspora erythraea*. J. Gen. Microbiol. 134:1773–1778.
94. Stuttard, C. 1979. Transduction of auxotrophic markers in a chloramphenicol-producing strain of *Streptomyces*. J. Gen. Microbiol. 110:479–482.
95. Suarez, J. E., J. L. Caso, A. Rodriguez, and C. Hardisson. 1984. Structural characteristics of the *Streptomyces* bacteriophage ϕC31. FEMS Microbiol. Lett. 22:1137–117.
96. Suarez, J. E., T. M. Clayton, A. Rodriguez, M. J. Bibb, and K. F. Chater. 1992. Global transcription pattern of ϕC31 after induction of a *Streptomyces coelicolor* lysogen at different growth stages. J. Gen. Microbiol. 138:2145–2157.
97. Sumby, P., and M. C. M. Smith. 2002. Genetics of the phage growth limitation (Pgl) system of *Streptomyces coelicolor* A3(2). Mol. Microbiol. 44:489–500.
98. Sumby, P., and M. C. M. Smith. 2003. Phase variation in the phage growth limitation system of *Streptomyces coelicolor* A3(2). J. Bacteriol. 185:4558–4563.
99. Susskind, M. M., and D. Botstein. 1978. Molecular genetics of bacteriophage P22. Microbiol. Rev. 42:385–413.
100. Van Mellaert, L., L. Mei, E. Lammertyn, S. Schacht, and J. Anne. 1998. Site-specific integration of bacteriophage VWB genome into *Streptomyces venezuelae* and construction of a VWB-based integrative vector. Microbiology 144:3351–3358.
101. Vats, S., C. Stuttard, and L. C. Vining. 1987. Transductional analysis of chloramphenicol biosynthesis genes in *Streptomyces venezuelae*. J. Bacteriol. 169:3809–3813.
102. Walter, F., M. Hartmann, and S. Klaus. 1981. Restriction endonuclease analysis of DNA from the *Streptomyces* phages SH3, SH5, SH10 and SH13. Gene 13:57–63.
103. Wellington, E. M. H., and I. K. Toth. 1994. Actinomycetes: Methods of Soil Analysis, part 2 Microbiological and Biochemical Properties, vol. 5. Soil Science Society of America, Madison, Wis.
104. Wikoff, W. R., L. Liljas, R. L. Duda, H. Tsuruta, R. W. Hendrix, and J. E. Johnson. 2000. Topologically linked protein rings in the bacteriophage HK97 capsid. Science 289:2129–2133.
105. Wilson, S. E., C. J. Ingham, I. S. Hunter, and M. C. M. Smith. 1995. Control of lytic development in the *Streptomyces* temperate phage ϕC31. Mol. Microbiol. 16:131–143.
106. Wilson, S. E., and M. C. M. Smith. 1998. Oligomeric properties arid DNA binding specificities of repressor isoforms from the *Streptomyces* bacteriophage ϕC31. Nucleic Acids Res. 26:2457–2463.

107. Zhou, X., Z. Deng, D. A. Hopwood, and T. Kieser. 1994. Characterization of ϕHau3, a broad-host-range temperate *Streptomyces* phage, and development of phasmids. J. Bacteriol. 176:2096–2099.

108. Zhou, X., Z. Deng, D. A. Hopwood, and T. Kieser. 1994. *Streptomyces lividans* 66 contains a gene for phage resistance which is similar to the phage lambda ea59 endonuclease gene. Mol. Microbiol. 12:789–797.

109. Ziegelin, G., E. Scherzinger, R. Lurz, and E. Lanka. 1993. Phage P4 alpha protein is multifunctional with origin recognition, helicase and primase activities. EMBO J. 12:3703–3708.

40

Mycoplasma Phages

JACK MANILOFF
KEVIN DYBVIG

Mycoplasma (with no italics) is the generic name for the small, wall-less bacteria that arose by degenerate evolution from the *Streptococcus* branch of the Gram-positive bacteria phylogenetic tree (25). At present these microorganisms are grouped into one putative genus plus eight well-accepted genera — including the *Acholeplasma*, *Spiroplasma*, and *Mycoplasma* — together forming the order Mollicutes (table 40-1). Mycoplasma genera have diverse habitats and biochemical and physiological properties. The genomes of some species are only 600–800 kbp.

The first mycoplasma phage was isolated in 1970 by R. N. Gourlay in England, using a bovine mycoplasma isolate as host cells and filtrates of bovine mycoplasma isolates as the phage source (16). Subsequently the host strain was identified as *Acholeplasma laidlawii* and the phage was identified as a filamentous phage containing single-stranded DNA. Over the next few years Gourlay isolated enveloped quasi-spherical phages as well as short-tailed phages infecting *A. laidlawii*. Since then, other workers have isolated more *Acholeplasma* phages as well as phages infecting *Spiroplasma* and *Mycoplasma* strains (23, 24). Phages active against other Mollicutes genera have not been isolated.

All mycoplasma phages that have been isolated are DNA phages (table 40-2). Mycoplasma phages with circular, single-stranded DNA genomes can be icosahedral or quasi-spherical in addition to filamentous. Most double-stranded DNA mycoplasma phages have short tails and linear DNA with particular terminal features, though an enveloped quasi-spherical double-stranded DNA phage has also been isolated. Mycoplasma phages with long tails have been reported but not been propagated. The absence of mycoplasma cell walls means mycoplasma phage adsorption must resemble that of animal viruses (i.e., adsorption to cell membranes) rather than that of bacteriophages (i.e., adsorption to cell walls or extracellular bacterial structures). In addition, since *Spiroplasma* and *Mycoplasma* genera evolved to use UGA as a tryptophan codon rather than a stop codon, *Spiroplasma*- and *Mycoplasma*-phage genomes must have coevolved with this codon change.

Data on many mycoplasma phages have been reviewed and these references should be consulted for the original literature citations (23, 24, 32). In reviewing the early mycoplasma literature, it should be noted that there have been significant changes in mycoplasma taxonomy and nomenclature over the years. In keeping with current virus taxonomy (47), mycoplasma viruses are now referred to as mycoplasma phages although mycoplasma phage particles continue to be called virions.

ssDNA Filamentous Phages

Filamentous single-stranded (ss) DNA *Acholeplasma* phages are short, rod-shaped particles. There have been more than 50 reports of their isolation but only one, designated *Acholeplasma* phage L51 and infecting some *A. laidlawii* strains, has been characterized. Filamentous *Spiroplasma* phages, by contrast, are longer and more filamentous than the equivalent *Acholeplasma* phage. About 60% of *Spiroplasma* cultures produce filamentous phages. The first isolated was from *Spiroplasma citri* and designated *Spiroplasma* phage SpV1.

Acholeplasma *Phage L51*

Virion and Macromolecules

L51 virions are non-enveloped, bullet-shaped particles, 14 nm by 71 nm, with one end rounded and the other irregularly shaped or flat. Virions have helical symmetry with a subunit spacing of 4.8 nm. As regards virus taxonomy, phage L51 is the type species of the genus *Plectrovirus* (small filamentous phages) in the family *Inoviridae* (filamentous phages containing ssDNA) (47) (see chapter 2 for a discussion of phage classification). L1 virions are resistant to treatment with the nonionic detergents Nonidet P-40 and Triton X-100; with the ionic detergent Sarkosyl NL97;

Table 40-1 Properties of the Genera in the Class Mollicutes[a]

Taxonomy	Oxygen requirement[b]	Sterol requirement	Genome size (kbp)	Habitat
Family: *Mycoplasmataceae*				
Genus: *Mycoplasma*	FA	Yes	580–1350	Animals and humans
Genus: *Ureaplasma*	FA	Yes	760–1170	Animals and humans
Family: *Entomoplasmataceae*				
Genus: *Entomoplasma*	FA	Yes	790–1140	Plants and insects
Genus: *Mesoplasma*	FA	No	870–1100	Plants and insects
Family: *Spiroplasmataceae*				
Genus: *Spiroplasma*	FA	Yes	780–2200	Plants and insects
Family: *Acholeplasmataceae*				
Genus: *Acholeplasma*	FA	Yes	1500–1650	Animals, some plants, and insects
Genus: Phytoplasma[c]	–	–	640–1185	Plants and insects
Family: *Anaeroplasmataceae*				
Genus: *Anaeroplasma*	OA	Yes	1500–1600	Bovine and ovine rumens
Genus: *Asteroleplasma*	OA	No	1500	Bovine and ovine rumens

[a] Based on data reviewed in (31).

[b] FA, facultative aerobe; OA, obligate anaerobe.

[c] Phytoplasmas have not been cultured and, therefore, have no official taxonomic status. These microorganisms may be obligate intracellular parasites. Phytoplasmas form a putative genus phylogenetically close to *Acholeplasma* (25). Although no phytoplasma growth data are available, genome sizes have been determined from phytoplasma-infected plant and insect tissues.

Table 40-2 Properties of characterized mycoplasma phages[a]

Nucleic acid	Morphology	Phage	Host	Genome size	Genome properties
ssDNA	Filamentous	L51	*A. laidlawii*	4.3–4.5 kb[b]	Circular
		SpV1	*S. citri*	6.8–8.3 kb[c]	Circular
	Icosahedral	SpV4	*S. melliferum*	4421 nt	Circular
	Enveloped, quasi-spherical	L172	*A. laidlawii*	14.0 kb	Circular
dsDNA	Short-tailed phage	L3	*A. laidlawii*	39.4 kbp	Linear, circular permuted, terminally redundant
		SpV3	*S. citri*	21.0 kbp	Linear, circular permuted, terminally redundant
		ai	*S. citri*	16.0 kbp	Linear, cohesive ends
		P1	*M. pulmonis*	11,660 bp	350 bp inverted terminal repeats, 5′-terminal proteins
	Enveloped, quasi-spherical	L2	*A. laidlawii*	11,965 bp	Circular
	Not determined	MAV1	*M. arthritidis*	15,644 bp	Linear

[a] References in the text.

[b] An L51-related strain, *Acholeplasma* phage L1, has a genome of 4491 nucleotides (M. Jaeger and G. Klotz, unpublished data, GenBank Accession No. X58839).

[c] As discussed in the text, three SpV1-related viruses have been sequenced, with genome sizes of 8273, 7768, and 6824 nucleotides.

with DNase I; and with pronase. Virions are also relatively heat and cold stable. On the other hand, virions are inactivated by UV irradiation with one-hit kinetics.

The L51 circular, ssDNA genome is 4.3–4.5 kb. An early transfer of Gourlay's original *Acholeplasma* phage isolate has been sequenced and shown to have a genome of 4491 nucleotides (nt) (M. Jaeger and G. Klotz, unpublished data, GenBank Accession No. X58839). Purified L51 virions, as analyzed by SDS-PAGE, contain four proteins of 70, 53, 30, and 19 kDa while the L51 genome sequence encodes four putative proteins of 31, 23, 19, and 12 kDa. The relationship between these putative L51 proteins and those observed from L51 virions is not known.

Growth and Replication

Phage L51 adsorption to *A. laidlawii* cells follows pseudo-first-order kinetics. The experimentally determined adsorption rate constant ($3–6 \times 10^{-9}$ cm^3/min at 37°C) is essentially the same as the theoretical value calculated for single-hit kinetics, so most phage–cell collisions result in adsorption. However, competitive adsorption studies show that only a small fraction of adsorbed L51 phages are bound to functional sites: of about 300 phages that can adsorb per colony forming unit (CFU), only 10–20 are bound to functional sites. There are limited data on the nature of *A. laidlawii* phage receptors, although a variety of cell membrane macromolecules have been proposed as receptors.

Penetration of phage L51 ssDNA appears to be coupled to its conversion to replicative-form (RF) double-stranded (ds) DNA. Studies of L51-infected *A. laidlawii* cells show the L51 DNA replication cycle, like that of ssDNA bacteriophages, proceeds in three steps: (i) synthesis of complementary-strand viral DNA by host cell gene products, converting parental ssDNA to RF dsDNA; (ii) RF replication by a rolling-circle mechanism requiring host cell and phage gene products to produce progeny RF dsDNA; and (iii) synthesis of progeny virus ssDNA from RF dsDNA by asymmetric DNA replication. An *A. laidlawii* REP^- mutant has been isolated that can propagate dsDNA *Acholeplasma* phages but not ssDNA *Acholeplasma* phages. In infections using these cells, L51 parental ssDNA is converted to intracellular RF dsDNA, but there is a block in RF replication and no progeny RF dsDNA is formed. An *A. laidlawii* 14 kDa protein is missing in REP^- cells and may be a cell gene product required for RF replication. Phage L51 assembly and release has been only minimally studied. An L51 DNA–protein complex has been identified late in infection. The complex contains L51 progeny ssDNA and L51 70 and 53 kDa structural proteins, but in different stoichiometric ratios than in L51 virions.

Growth studies originally indicated that L51 phage has a noncytocidal infectious cycle leading to persistently infected cells. In agreement with this model, there is no loss of viability of L51-infected *A. laidlawii* cells up to an MOI of 10. However, infected cells grow slower and make smaller colonies than uninfected cells, which probably explains the turbid plaques formed by L51 phage. Data on the properties of *A. laidlawii* cultures also indicate that *A. laidlawii* strains persistently infected with L51-related phages are prevalent in nature.

One-step growth curves of L51-infected *A. laidlawii* cells show a 10–15 minute latent period followed by increasing plaque-forming units (PFU) over the next 2–3 hours (the rise period), with a yield of 150–200 progeny phage per infectious center at 2–3 hours post-adsorption. The latent and eclipse periods are indistinguishable in artificial-lysis experiments, indicating progeny phage assembly and release are coupled. The rate of virus release decreases 2–3 hours post-adsorption but the progeny phage titer continues to increase slowly, with no measurable loss in CFU. These growth studies were the first indication that L51 has a noncytocidal infectious cycle. See chapter 5 for a general discussion of phage life-cycle characters of infection and release.

Spiroplasma *Phage SpV1*

Virion and Macromolecules

As regards virus taxonomy, SpV1-related phages are classified as species in the genus *Plectrovirus* in the Family *Inoviridae* (47). This classification is clearly wrong because SpV1-related phages are significantly different from the *Plectrovirus* type species, *Acholeplasma* phage L51 (above), in both virion morphology and genome. It can be expected that SpV1-related phages will be reclassified, probably as a separate genus in the family *Inoviridae.* Although several SpV1-related phage isolates (e.g., SpV1-R8A2B, SpV1-C74, and SVTS2) have been studied, the available data on the basic virology of SpV1-related phages are still limited.

SpV1 virions are long, non-enveloped, filamentous particles, 10–15 nm by 230–280 nm, with one end rounded and a flat plate at the other (23, 24). Micrographs also show filamentous particles, 10–15 nm in diameter, with lengths 2 or more times longer than virions. SpV1 virions have a density of 1.39 g/cm^3 in CsCl and 1.21 g/cm^3 in metrizamide. SpV1 biological activity is resistant to the nonionic detergents Nonidet P-40 and Triton X-100; sensitive to chloroform and ether; and heat stable. The genomes of three SpV1-related phages have been sequenced and analyzed: SpV1-R8A2B, SpV1-C74, and SVTS2.

Phage SpV1-R8A2B has a circular, ssDNA genome of 8273 nucleotides (nt) with a 22.9% G + C content, somewhat less than the 26% G + C content of its *S. citri* host cells (32). From the sequence, 14 open reading frames (ORFs) were identified on the viral strand (the strand in the virion, which is the + strand) and four putative ORFs were identified on the complementary strand. One of the putative viral-strand gene products has significant similarity to transposases of the insertion sequence 30 (IS*30*) family (29).

Phage SpV1-C74 has a circular, ssDNA genome of 7768 nt, with 13 putative ORFs encoded in the viral strand and two in the complementary strand (32). Eleven viral-strand ORFs and two complementary-strand ORFs are conserved between the SpV1-R8A2B (above) and SpV1-C74 genomes. The putative SpV1-R8A2B IS*30* transposase ORF is not found in the SpV1-C74 genome. However, one of the SpV1-C74 ORFs, not found in the SpV1-R8A2B genome, has significant sequence similarity to transposases of the insertion sequence 3 (IS*3*) family (29).

Phage SVTS2 has a circular, ssDNA genome of 6824 nt, with a G + C content of 22.7% (38). The viral strand encodes 14 putative ORFs. Nine of these have sequence similarity to ORFs found in both phages SpV1-R8A2B and SpV1-C74. The SVTS2 genome lacks an ORF with recognizable similarity to a transposase. There is also sequence similarity between an ORF encoding a gene product of unknown function identified in all three SpV1-related phages and an ORF in *Acholeplasma* phage L1.

Growth and Replication

There are limited data on SpV1 phage growth (23, 24). One-step growth curves of SpV1-infected *S. citri* host cells show a 1 hour latent period followed by increasing PFU over at least the next 6 hour. This suggests *Spiroplasma* phage SpV1 has a noncytocidal infectious cycle similar to those of the filamentous ssDNA *Acholeplasma* phage L51 and the ssDNA isosahedral *Spiroplasma* phage SpV4.

Multiple copies of full-length and partial genomes of SpV1 and SpV1-related phages have been found in strains of most *Spiroplasma* species examined (2, 32). The presence of phage-encoded transposases and inverted repeats at the termini of phage fragments integrated in cell chromosomes suggests that RF dsDNAs of phages SpV1-R8A2B and SpV1-C74 function as insertion elements of the IS*30* and IS3 families, respectively (2, 29). The extent of integrated SpV1-related phage sequences implies these phages may have a significant effect on *Spiroplasma* chromosome evolution.

Following phage SVTS2 infection of *S. citri*, resistant cells were selected and found to contain a 2.1 kbp fragment of the SVTS2 genome integrated into chromosomal and extrachromosomal DNA. Transfection of this fragment into phage SVTS2-sensitive *S. citri* transformed the host cell phenotype to SVTS2-resistance. This resistance phenotype appears to be due to inhibition of intracellular phage DNA replication rather than to a change in cell surface phage receptors (37).

ssDNA Icosahedral Phage

Spiroplasma *Phage SpV4*

Virion and Macromolecules

Spiroplasma phage SpV4 was isolated from and propagated in *Spiroplasma melliferum* strains (23, 32). Virions are icosahedral (T=1) particles that are 27 nm in diameter with 20 protrusions, each 5.4 nm long, making SpV4 the only known icosahedral mycoplasma phage (7). As regards virus taxonomy, *Spiroplasma* phage SpV4 is the type species of the genus *Spiromicrovirus* in the family *Microviridae* (small icosahedral phages containing ssDNA) (47).

The family *Microviridae* contains four genera: *Microvirus* (host: *Enterobacteriaceae*), *Spiromicrovirus* (host: *Spiroplasma*), *Bdellomicrovirus* (host: *Bdellovibrio*), and *Chlamydiamicrovirus* (host: *Chlamydia*). It has been proposed recently that these four genera form two subfamilies within the *Microviridae*, one subfamily consisting of the genus *Microvirus* and the other subfamily containing of the other three genera (4). Three shared features of this latter proposed subfamily are seen with the *Spiroplasma* phage SpV4: (i) the capsid lacks spikes, (ii) external head-scaffolding proteins are also lacking (4), and (iii) the major SpV4 coat protein is more complex than that of *Microvirus* phage (e.g., ϕX174). In particular, the SpV4 coat protein contains an insertion loop that forms the 5.4 nm protrusions that may function similar to *Microvirus* spike proteins (7).

SpV4 virions have a density of 1.40 g/cm^3 in CsCl and 1.24 g/cm^3 in metrizamide. SpV4 biological activity is resistant to the nonionic detergent Triton X-100, the ionic detergent sodium dodecyl sulfate, chloroform, ether, DNase, RNase, and proteinase K. The SpV4 virion nevertheless is more heat-sensitive than the ssDNA filamentous phages infecting the *Acholeplasma*.

Spiroplasma phage SpV4 has a circular, ssDNA genome of 4421 nt with a G-C content of 32%—higher than its host *S. melliferum's* 26% G-C content (32). A total of nine ORFs are encoded, all on the viral strand (the strand in the virion that is also the + strand). ORF1 encodes the 64.0 kDa major capsid protein (corresponding to the 60 kDa virion protein identified by SDS-PAGE) which displays significant sequence and structural similarity to the major capsid proteins of *Bdellomicrovirus* and *Chlamydiamicrovirus* phages. Two other SpV4 ORFs have significant similarities to *Chlamydiamicrovirus* ORFs, one of which is a putative endonuclease. The nine SpV4 ORFs would produce proteins of 64.0, 32.1, 17.3, 14.1, 9.5, 8.5, 5.6, 4.9, and 3.8 kDa. Only the 64.0 kDa protein, the major capsid protein, has been identified in virions or in SpV4-infected cells.

Growth and Replication

One-step growth curves of SpV4-infected *S. melliferum* show a 1–2 hour latent period followed by a 4 hour period of rapid progeny phage release. The progeny phage yield gradually reaches a plateau at 20 hour post-adsorption of 100–200 phage progeny per infectious center. These results, together with the fact that SpV4 produces clear plaques, indicate that SpV4 has a nonlytic, cytocidal productive infection cycle in which infected cells are no longer viable but continue to release progeny virions for many hours. This type of infectious cycle is different from the lytic infections of phages in the other *Microviridae* genera, but similar to that found for the short-tailed mycoplasma phages (described below).

Sequence analysis of the SpV4 genome indicates that DNA replication, transcription, and translation are similar to those of other ssDNA phages. An SpV4 intergenic inverted repeat sequence, with seven GC base pairs, may form the hairpin structure used by RNA polymerase or primase for synthesis of the RNA primer for complementary strand synthesis as in other ssDNA phages. Each of the nine SpV4 ORFs has an upstream Shine–Dalgarno sequence, and eight ORFs start with an ATG codon and one with a GTG codon. Transcription appears to use promoter and rho-independent termination sites similar to those in Gram-positive bacteria.

Four SpV4 mRNAs (with sizes 2.7, 3.4, 4.4, and 7.8 kDa) have been identified in SpV4-infected host cells at all times post-adsorption. The latter mRNA is longer than one genome in length. It has been proposed that SpV4 transcription starts at different promoters and stops at the single rho-independent termination site identified by genome sequence analysis. In this model, the 3.4 and 7.8 kDa mRNAs start at the same promoter, but synthesis of the 7.8 kDa mRNA reads through the termination site, continues and transcribes the entire genome, and stops at the termination site when it reaches it again.

dsDNA Short-Tailed Phages

One of Gourlay's original *Acholeplasma* phage isolates was a virus morphologically similar to coliphage T7 (see chapter 20): short-tailed phages containing linear dsDNA. This was a surprising morphology for a mycoplasma phage because mycoplasma cells have no cell wall and it was not expected that phage tails would have a role in adsorption in such situations. This original short-tailed mycoplasma phage was designated L3. Another, apparently similar short-tailed *Acholeplasma* phage has been reported but only L3 has been characterized (23, 24).

Short-tailed phages have also been found in over 60% of *Spiroplasma* strains examined, representing several *Spiroplasma* species. Based on genome size and structure (table 40-2), there are probably three types of short-tailed *Spiroplasma* phages: SpV3, *ai*, and SRO phages. Only limited data are available on these phages (23, 24).

Over 100 *Mycoplasma* species have been recognized (31). For the vast majority of these species, no evidence of phage exists. The relative lack of known *Mycoplasma* phages is hardly surprising given that most investigators do not design experiments with the goal of identifying novel phages. Also, the fastidiousness of mycoplasmas can render the original isolation and subsequent characterization of phages technically challenging. The only *Mycoplasma* phage that has been characterized is P1, a short-tailed dsDNA phage isolated from the murine pathogen, *Mycoplasma pulmonis*.

Acholeplasma Phage L3

Virion and Macromolecules

L3 virions have a polyhedral head with a diameter about 60 nm, a collar about 8 nm thick and 16 nm wide, a short tail about 10 nm wide and 20 nm long, and fibers attached to the collar. As regards virus taxonomy, *Acholeplasma* phage L3 is an unassigned species in the family *Podoviridae* (short-tailed phages) (47). L3 virions have a density of 1.477 g/cm^3 in CsCl, and are resistant to treatment with the nonionic detergents Nonidet P-40 and Triton X-100, and are also resistant to ether. Virions are inactivated by UV irradiation with one-hit kinetics and are relatively heat-sensitive.

Electron microscopic measurements found L3 virions contain linear dsDNA with a size of 39.4 kbp. Subsequent studies showed that this 39.4 kbp DNA consists of a 36.2 kbp L3 genome plus about 8% terminal redundancy. L3 DNA also has limited circular permutation, producing a circular 36.2 kbp restriction endonuclease map and suggesting that assembly may involve packaging from *pac* sites. Purified L3 virions contain 19 proteins as determined by SDS-PAGE. The major protein is 43 kDa and may be the major capsid protein. Complementation tests using L3 temperature-sensitive mutants have shown at least 21 complementation groups.

An interesting aspect of the L3 genome is that it contains no GATC sequences although it would be expected to have about 150 GATC sites. The absence of GATC sites was first suggested by the observation that L3 phages are not restricted by an *A. laidlawii* strain that restricts phages with DNA containing GATC sequences, and confirmed by studies showing that L3 DNA is resistant to restriction endonucleases that cleave unmethylated and methylated GATC sites. Hence, like a number of other phages, L3 has evolved under selective pressure for the loss of sequences recognized by host cell restriction endonucleases.

Growth and Replication

L3 adsorption to *A. laidlawii* cells follows pseudo-first-order kinetics with a rate constant of about 3×10^{-10} cm^3/min at 37°C. The rate constant can vary by a factor of 5 depending on the cells and media. The theoretical value calculated for single-hit kinetics is 2×10^{-9} cm^3/min. Therefore, depending on the experimental conditions, anywhere from a few percent to most phage–cell collisions result in adsorption. L3 adsorption requires Ca^{2+}, which cannot be replaced by Mg^{2+} or monovalent cations, and has been proposed to involve electrostatic interactions between virion and cell-membrane proteins.

Adsorbed L3 virions behave like polyvalent ligands: they diffuse along the cell membrane and crosslink receptors. It has been proposed that the fibers attached to the L3 virion collar function as polyvalent determinants. As adsorbed L3 virions diffuse on the cell membrane, the fibers crosslink mobile receptors producing capping or clustering of adsorbed virions. This capping process is dependent on energy metabolism by the adsorbed cell. Adsorbed L3 virions can also crosslink receptors on different cells, promoting cell fusion to produce the giant cells seen in L3-infected cultures.

The rate of total DNA synthesis in L3-infected cells decreases post-adsorption, but continues at a measurable rate for at least 17 hours. By 1 hour post-adsorption, about 40% of nascent DNA is phage DNA; by 2 hours, about 60–80% of nascent DNA is phage DNA; and by 3 hours essentially all nascent DNA is phage DNA. L3 must not obtain precursors for progeny DNA synthesis by re-utilizing host cell DNA nucleotides because: (i) although the host cell chromosome is progressively unfolded and possibly fragmented during L3 infection, no significant amount of acid-soluble cell DNA is released; (ii) no significant amount of cell DNA is recovered in L3 progeny; and (iii) if cell DNA nucleotides were the only L3 DNA precursors, then only about 100 progeny L3 could be produced, although many L3-infected cells produce more than 500 progeny phages. The small genome of L3 phage argues against it encoding gene products for de novo nucleotide synthesis, like those found in large genome phages. Instead, L3 DNA precursors must come from the salvage pathways used by mycoplasmas for nucleotide synthesis, in which medium-supplied free bases and nucleosides are converted to nucleotides. Salvage-pathway enzymes may also be encoded by L3 phage.

Intracellular L3 DNA replication involves fast-sedimenting complexes, indicating formation of concatemers. These may be formed by recombination between the terminally redundant ends of linear L3 DNA molecules. As for DNA synthesis, the rates of total RNA and protein synthesis in L3-infected cells decrease post-adsorption, but both continue at measurable rates for at least 17 hours. Approximately 20 phage-specific proteins (including 10 virion proteins) have been identified in L3-infected cells by SDS-PAGE. However, the long labeling times needed for these studies have precluded studies of the temporal regulation of L3 gene expression. L3 DNA packaging is probably from concatemers via a *pac*-site cutting mechanism that generates linear progeny phage DNAs with terminal redundancy and limited circular permutation.

Growth of L3-infected cells stops at the start of infection. One-step growth curves of L3 infected *A. laidlawii* show a 90 minute latent period followed by a linear increase in progeny phage that continues for about 15 hours. This also was shown by "single burst" experiments in which the average phage yield per infected cell was found to increase with time post-adsorption, and by measurements of phage production by individual cells as a function of time post-adsorption. The latter studies found that every L3-infected cell that begins to release progeny phage during the first few hours post-adsorption continues to do so for at least 24 hours post-adsorption. Artificial lysis experiments show a 60 minute eclipse period followed by intracellular accumulation of infectious progeny phage. Since L3 produces clear plaques, these data suggest that *Acholeplasma* phage L3 produces a cytocidal infection with progeny phage release by a nonlytic mechanism.

By 8 hours post-adsorption, the cytoplasmic face of the host membrane is covered by L3 virions, oriented radially with tails facing the membrane. Extracellular membrane vesicles enclosing one or more L3 progeny virions are seen in L3-infected cell preparations. It is not known whether these vesicles are a budding mechanism for progeny L3 virions release or whether there is some other means for the release of non-enveloped L3 virions. L3 phage-containing vesicles somehow must bud from infected cells without destroying cell integrity, since infected cells continue to release progeny phage for many hours. The vesicles may eventually break down to release non-enveloped L3 virions.

With the exception noted below, L3 is the only *Acholeplasma* phage reported, thus far, that is able to propagate on an *Acholeplasma* species other than *A. laidlawii*. L3 forms plaques on lawns of *A. laidlawii*, *A. modicum*, and *A. oculi*. The exception noted above is recent studies which indicate that *Acholeplasma* phage L2 can be propagated on *A. laidlawii* and *A. oculi* strain ISM1499 (K. Dybvig, unpublished data).

Spiroplasma *Phage SpV3*

Virion and Macromolecules

SpV3 has been isolated from *S. citri* and *S. mirum* strains. Virions have a polyhedral head, about 40 nm in diameter, and a short tail, 13–8 nm by 6–8 nm. As regards virus taxonomy, *Spiroplasma* phage SpV3 is an unassigned species in the Family *Podoviridae*, where it is listed as phage C3 (an earlier designation) (47). SpV3 virions have been reported to have a density of 1.45 g/cm^3 in CsCl. However, since it has also been reported that banding in CsCl causes partial SpV3 dissociation and DNA release, the density data must be questioned. The SpV3 genome is linear dsDNA, 21 kbp in size with about 5% terminal redundancy and circular permutation. Preliminary data indicate the SpV3 G-C content is 27%, experimentally the same as the 26% G+C content of *S. citri* DNA. Different numbers and sizes of proteins have been reported for SpV3. It is not known whether this is due to differences between different isolates or experimental differences between different laboratories studying the same phage.

Growth and Replication

Electron microscopic studies show accumulations of intracellular progeny virions and membrane-bounded budding phages. This suggests that release of progeny phage from SpV3-infected cells involves budding similar to the type of release described above for *Acholeplasma* phage L3. No additional data are available on SpV3 replication.

Spiroplasma *Phage* ai

Virion and Macromolecules

Phage *ai* has been isolated from *S. citri* and has a morphology similar to that of phage SpV3: polyhedral heads, 43–54 nm in diameter, and 14 nm long tails. Phage *ai* as well as SRO phages (discussed below) have not been formally classified but presumably are members of family *Podoviridae*. Phage *ai* virions have a density of 1.45 g/cm^3 in CsCl and are resistant to treatment with the nonionic detergent Triton X-100. The phage *ai* genome is linear DNA of 16 kbp with cohesive ends. Linear, circular, and concatemeric *ai* DNA has been identified in *ai*-infected cells. Heating and S1 nuclease treatments convert the circular and concatemeric DNA to linear 16 kbp DNA. The virions of *ai* and two *ai*-related phages have seven proteins of sizes 86, 65, 63, 55, 47, 45, and 26 kDa.

Growth and Replication

Phage *ai* seems to have a nonlytic, cytocidal productive infection cycle similar to that described above for *Acholeplasma* phage L3 but, unlike L3, phage *ai* can also have a lysogenic cycle. One-step growth and artificial lysis experiments show the *ai* productive infection cycle has an eclipse period of about 3 hours, after which time intracellular accumulation of progeny virions begins. The *ai* latent period is 5–6 hours, after which time progeny-phage release begins. Progeny phage release continues until about 11 hours post-infection, yielding 40 progeny *ai* per infected host.

All *S. citri* strains that have been examined contain crytpic *ai* prophage consisting of less than full-length *ai* DNA fragments integrated in the cell chromosome. Some *ai*-infected cells produce *ai*-lysogens with an *ai* genome integrated in the cell chromosome within a cryptic *ai* prophage, presumably by homologous recombination. Such lysogens are immune to superinfection by homologous phages but can be infected by heterologous phages. Although some *ai* phage is spontaneously released from lysogens, the lysogenic phenotype is retained after repeated lysogen passage in media containing antiserum to *ai* phage. Hence, the lysogenic phenotype appears not to be due to some type of persistent infection. Lysogens do not adsorb *ai* phage. This indicates that immunity to superinfection in *ai* lysogens is similar to that for *Salmonella* phage lysogeny, where immunity involves modification of cell-surface phage receptors.

Spiroplasma *SRO Phages*

SRO (sex ratio organisms) are spiroplasmas that infect some *Drosophila* and affect the sex ratio of their progeny. These SRO spiroplasmas are infected by short-tailed phages. Unfortunately, the few studies of SRO phages were done before a medium for growing SRO was developed, so the data are limited to preparations isolated from SRO-infected *Drosophila* (23).

Virion and Macromolecules

Several SRO phage strains have been studied. All have a short-tailed phage morphology with a polyhedral head, 35–45 nm in diameter, a short tail, 10–12 nm by 7–9 nm, and a baseplate. These dimensions are similar to those of *Spiroplasma* phages SpV3 and *ai*. SRO phage preparations have been reported to contain linear dsDNAs of 17, 21.8, and >30 kbp. In some cases, more than one of these DNAs have been found in the same SRO phage preparation. Restriction endonuclease cleavage and hybridization studies of these DNAs have produced confusing results, as expected for studies of phages that have not been cloned or purified.

Growth and Replication

Like SpV3, electron microscopic studies of phage-infected SRO cells show accumulations of intracellular progeny virions and membrane-bounded budding phages, suggesting that progeny SRO phages are released from infected cells by budding.

Injection of SRO spiroplasmas from one *Drosophila* species into another SRO spiroplasma-infected *Drosophila* species can produce interference with the sex ratio trait, in some cases leading to loss of the trait and return to a normal progeny sex ratio in the recipient *Drosophila* or in their daughters. This loss of pathogenicity appears to be caused by clumping of the two SRO spiroplasmas and eventual lysis of the SRO in the recipient *Drosophila*. Elimination of the sex ratio trait has been shown to be phage-mediated. Preliminary studies have shown that only SRO phage-infected spiroplasma cells produce this effect and that the SRO phage determines the clumping specificity of its host cells. SRO phage-infected spiroplasmas only clump with spiroplasmas from a different *Drosophila* species. This phenotype change in SRO phage-infected spiroplasmas may be similar to the lysogenic conversion and modification of the host cell surface described above for Spiroplasma phage *ai*.

Mycoplasma *Phage P1*

Virion and Macromolecules

P1 virions have a short tail and an isometric head only 28 nm in diameter (11). As regards virus taxonomy, *Mycoplasma* phage P1 is presumably a member of the family *Podoviridae* (47), although it has not been formally classified. P1 infectivity is resistant to treatment with the nonionic detergent Triton-X 100, chloroform, DNase I, and RNase A. Infectivity is lost by treatment with sodium dodecyl sulfate, deoxycholate, and proteinase K (11).

P1 virions contain a linear, dsDNA genome of 11,660 bp with inverted terminal repeats of 350 bp (46, 52). The P1 DNA G-C content is 26.8% G-C, experimentally indistinguishable from its host's 26.6% G-C content (6). From the complete nucleotide sequence, 11 putative P1 genes were identified (46). These are organized such that transcription can be initiated within each inverted terminal repeat and proceed toward the middle of the genome, with six genes being transcribed from one terminus and five from the other. Consistent with this gene organization, putative promoters oriented inward were identified within the inverted terminal repeats and a transcription terminator was identified in the middle of the genome where the two opposing RNA polymerase molecules would meet.

Considerable evidence indicates a terminal protein is attached to the 5′-terminus of each P1 DNA strand (52). Initial attempts to identify P1 DNA by agarose gel electrophoresis of nucleic acids extracted from P1 virions were unsuccessful. Pretreatment with proteinase K resulted in P1 DNA that migrated in agarose gels as linear molecules. P1 DNA was resistant to digestion with phage λ exonuclease, an enzyme that specifically degrades from DNA 5′-termini. Electron microscopic analysis showed the presence of globular material at the ends of the P1 genome. The globular material, presumably protein, was absent on P1 DNA that had been treated with proteinase K. P1 genome termini also have a low level of nucleotide sequence similarity to the termini of the genome of phage φ29, a well-characterized phage that has terminal proteins attached to the ends of its linear dsDNA genome (36) (see chapter 22).

P1 virion proteins have not been studied. Tentative assignments have been suggested for several proteins predicted from the P1 genome sequence (46). The P1 ORF8 gene product has a collagen-like repetitive motif similar to that of some bacteriophage tail fiber proteins (44). P1 ORF1 is predicted to encode a phage DNA replication protein. The putative product of P1 ORF3 has a low level of amino acid sequence similarity to the terminal protein of φ29 and related phages, but experimental verification is needed to conclude that P1 ORF3 actually encodes a terminal protein. The other ORFs in the P1 genome encode proteins with no significant sequence similarity to proteins in sequence databases.

Growth and Replication

The pseudo-first-order rate constant for P1 adsorption is very low, 2×10^{-11} cm^3/min at 37°C. P1 infection is cytocidal. Host variants to which P1 fails to adsorb can be isolated at a high frequency (10). *M. pulmonis* cells produce a family of phase-variable lipoproteins referred to as V-1 antigens or Vsa proteins (3, 39). Each cell can produce only one Vsa protein because the genome has a single *vsa* expression site that promotes transcription and translation of the gene that occupies this site (3, 6, 39). Alternative *vsa* genes can become expressed as a result of site-specific DNA inversions that bring a gene into the expression site while moving the previously expressed gene to a silent site. P1 virus adsorbs to cells that produce the VsaA protein. Cells that produce VsaB are resistant to infection due to failure of the virus to adsorb (10). Cells that produce VsaC and VsaE proteins are also resistant to P1 infection, presumably due to lack of P1 adsorption (19). Because P1 virus can infect only cells that produce VsaA, VsaA is a candidate for being the virus receptor. All Vsa proteins have an identical N-terminal domain of 242 amino acids. The unique feature of the VsaA protein is a 17 amino acid sequence that is repeated in tandem about 40 times (3). This repetitive sequence may interact with P1 virions, possibly through the collagen-like repeat of the P1 ORF8 gene product.

P1 was originally observed as phage-like particles in electron micrographs of *M. pulmonis* strain 5884 (11). Several, but not all, strains of *M. pulmonis* can serve as phage indicators, with strain UAB 6510 and its derivatives being used for P1 isolation and characterization (11). Plaques are unusually heterogeneous in size and turbidity, an observation that can be explained by high-frequency changes in the host. Variations in *M. pulmonis* Vsa surface proteins affect the ability of P1 to adsorb to host cells (12, 19), and variations in the production of phase-variable restriction-modification systems affect the ability of progeny phage to successfully infect neighboring cells in a developing plaque (12). Thus, the dynamics of interactions between P1 phage and host cells are complex. One-step growth curves have a latent period of 1.5 hours and a rise period over 4 hours (11). By 6 hours post-infection, 100–500 progeny phage have been released per infectious center.

The terminal protein probably serves as the primer for P1 DNA replication as it does for other phages with terminal proteins (35). However, intracellular replication, assembly, and release of P1 virions have not been studied. Numerous virus-like particles that appeared to be intracellular and in close proximity to the cytoplasmic membrane were noted in the initial electron micrographs of *M. pulmonis* strain 5884 (R. Miles, personal communication). Therefore, in contrast to the short-tailed phages of *Acholeplasma* (23, 24), P1 virions may be released from infected cells in a single burst as opposed to budding in a non-lytic manner.

dsDNA Quasi-spherical Phages

One of Gourlay's original *Acholeplasma* phage isolates was an enveloped quasi-spherical virion containing dsDNA and infecting some *A. laidlawii* strains. This isolate was designated *Acholeplasma* phage L2 and, although other apparently similar isolates have been reported, only L2 has been studied in detail (23, 24, 28).

Acholeplasma *Phage L2*

Virion and Macromolecules

Acholeplasma phage L2 has a unique morphology: virions are enveloped, quasi-spherical particles, 50–125 nm in diameter. The broad virion size range is due to heterogeneity of L2 preparations. Infection by an L2 virion results in the production of three morphological forms of progeny phages, designated L2-I, L2-II, and L2-III. These are produced in about the same relative amounts — the ratio of L2-I to L2-II to L2-III is 4–16 to 2–3 to 1 — regardless of which form is used to start an infection. The virion diameters of L2-I, L2-II and L2-III are 74, 88, and 132 nm, respectively, based on measurements in aqueous media that yield slightly larger sizes than electron microscopic measurements.

All three L2 forms contain unit-length DNA molecules, so the larger forms are not the result of concatemer packaging. However, UV inactivation studies indicate that, while L2-I and L2-III contain one genome copy per virion, L2-II contains two to three genome copies per virion. Electron micrographs show virions have a quasi-spherical, densely stained core bounded by a membrane. These data and the absence of any defined capsid structure have led to the proposal that L2 virions are nucleoprotein condensations within a lipid–protein membrane. The lipid composition and that of the *A. laidlawii* cell membranes have been studied in several laboratories, but there is no consensus on the relative amounts of the major lipid classes (glycolipids, phospholipids, phosphoglycolipids, and neutral lipids). L2 virion and host cell membranes have been shown to have essentially identical fatty acid compositions.

As regards virus taxonomy, *Acholeplasma* phage L2 is the type species (and only classified member) of the Genus *Plasmavirus* in the family *Plasmaviridae* (enveloped quasi-spherical phages containing dsDNA) (47). Efforts to measure L2 density have been confounded by dehydration of virions within the sucrose gradients used for the experiments, and a true equilibrium density has yet to be determined. L2 biological activity is sensitive to treatment with nonionic detergents Brij-58, Triton X-100, and Nonidet P-40; sensitive to ether; and sensitive to chloroform. Virions are also sensitive to treatment with pronase and trypsin, but not to DNase I or to phospholipase A_2. L2 virions are extremely heat-sensitive, but relatively cold-stable.

L2 virions contain circular dsDNA with a genome size of 11,965 bp and 32.0% G-C content, which is experimentally indistinguishable from the *A. laidlawii* 31.8% G-C (28). Sequence analysis has identified 15 putative ORFs, all on one DNA strand, designated ORF1 to ORF14 and ORF13*. The ORFs are clustered in four groups separated by noncoding intergenic regions. Sequence analysis of the L2 genome indicates that transcription uses promoter- and rho-independent termination sites, similar to those in Gram-positive bacteria, to produce polycistronic mRNAs of gene clusters. Several possible cases of transcriptional read-through have also been identified. Each of the 15 L2 ORFs has an upstream Shine–Dalgarno sequence and starts with an ATG codon. Translational coupling or reinitiation may be involved in translation of the polycistronic mRNAs.

The gene product of ORF 5, the only L2 ORF with significant sequence similarity to any database sequence, is the putative L2 integrase based on its sequence similarity to site-specific DNA recombinases. ORF13 and ORF14 contain N-terminal signal sequences of 27 and 26 amino acids, respectively, indicating that their putative gene products may be integral membrane proteins. ORF13* has its start codon within and in the same reading frame as ORF13 and, therefore, the ORF13* gene product must be the 443 amino acid C-terminal fragment of the 738 amino acid ORF13 gene product. The putative ORF12 gene product is a 17 kDa basic protein, which may be the major virion 19 kDa protein identified by SDS-PAGE. In agreement with the suggestion that the 19 kDa protein is a DNA-binding protein, the ORF12 putative gene product is the only L2 gene product with extensive helix–turn–helix motifs, characteristic of DNA-binding proteins.

Analysis of purified L2 virions by SDS-PAGE shows four proteins: 64, 61, 58, and 19 kDa. The 64 kDa protein appears to be an L2 integral membrane protein based on its reaction with a variety of reagents and may correspond to the ORF13 gene product, which would be a 78 kDa protein after cleavage of its signal sequence. The 61 kDa protein appears to be a peripheral membrane protein, perhaps corresponding to the ORF1 gene product, which would be a 67 kDa protein. The 19 kDa protein is the major L2 protein and does not appear to be a membrane protein. It has been proposed to be a DNA-binding protein involved in nucleoprotein core assembly and may correspond to the ORF12 gene product, which would be a 17 kDa protein with extensive helix–turn–helix motifs.

Growth and Replication

L2 adsorption to *A. laidlawii* cells follows pseudo-first-order kinetics with an adsorption rate constant, depending on the medium being used, that is 10–100 times less than the theoretical value calculated for single-hit kinetics. Therefore, depending on the medium, only 0.1–10% of phage–cell collisions result in L2 adsorption. This may be due to a limiting number of L2 receptors or to a rate-limiting step in the interaction of virion and cell-membranes. Studies of the molecular nature of L2 receptors have produced conflicting data, so there are no conclusions on which types of cell-membrane macromolecules are involved in L2 adsorption. It is assumed that adsorption leads to fusion of L2 phage and host cell membranes, resulting in entry of the virion nucleoprotein condensation into the cell and then its uncoating. Note that recent studies indicate that *Acholeplasma* phage L2 can also be propagated on *A. oculi* strain ISM1499 (K. Dybvig, unpublished data) suggesting a co-occurrence of the phage L2 cell-surface receptor in *A. laidlawi*, phage L2's normal host, and the *A. oculi* strain.

L2 one-step growth curves have a 1–2 hour latent period followed by a gradual rise period that levels off at 6–10 hours post-infection. Artificial lysis experiments show that the latent and eclipse period are the same, indicating that L2 assembly and release are coupled and that L2 infection is noncytocidal. By 10 hours post-infection, 100–1000 progeny phage have been released per infected cell, depending on the host cell strain being used. Since there is no measurable loss in cell titer in L2-infected cultures at an MOI up to 30 and since maturation by budding from infected host cell membranes is shown by electron microscopy (hence the similarity of virion and cell fatty acid composition reported above), most if not all infected cells must be noncytocidally infected.

Pulse-labeling and restriction endonuclease fragment analysis have shown that L2 DNA replicates bidirectionally from two *ori* sites. These sites have been mapped and sequence analysis has located each *ori* site in an intergenic region of the L2 genome. Each putative *ori* site contains a Dna-A box bounded by AT-rich 6-mer repeats. Intracellular L2 DNA replication is membrane-associated and appears to involve a host DNA polymerase similar to Gram-positive bacterial DNA polymerase III. Continuous- and pulse-labeling studies also show that L2 DNA replication in infected cells continues throughout the growth-curve rise period. L2 DNA replication and progeny virion maturation continue for several hours after integration of a phage genome into the host cell genome. Although L2 progeny DNA replication stops about 5–6 hours post-infection, cytoplasmic progeny L2 DNA persists up to at least 10 hours post-infection.

Lysogeny

Phage L2 has a unique type of infection cycle, with productive noncytocidal replication (above) that is followed by lysogeny in most if not all L2-infected cells. About 2–4 hours post-infection, L2 infection of *A. laidlawii* leads to establishment of lysogeny with the circular L2 genome integrated at a unique site in the host cell chromosome. Based on extensive similarity to the *attP* site of phage λ, the L2 *attP* integration site was mapped to 280 bp downstream from the putative L2 integrase gene (ORF5), within a 600 bp intergenic region. This is similar to the situation in other temperate phages with lysogeny involving site-specific integrases, in which *attP* sites are downstream of integrase genes (see chapter 7 for a review of phage integration). The L2 *attP* site is a 25 bp sequence with a 9 bp inverted repeat: CATCTTCAT–7 bp–CTGAAGATA.

L2 lysogens are immune to superinfection by homologous, but not by heterologous, phages. Two models of immunity exist, both examples of lysogenic conversion: (i) expression of an L2 repressor which inhibits intracellular L2 gene expression (and thereby productive infection) and (ii) modification of the cell surface, thereby affecting phage adsorption. Since L2 lysogens also are immune to superinfection by transfecting L2 DNA, immunity seems at least due to an L2 repressor. By contrast, phage *ai* lysogens appear to be immune at least due to modification of cell surface receptors. Consistent with this repression model of L2 immunity, lyosgens can be induced to produce L2 phages by mitomycin C and by UV treatment (see chapter 8). Though not necessarily leading to cell-surface-mediated immunity, during productive infection the osmotic stability of L2-infected cells increases, which is perhaps due to an increase in cell permeability during the period of L2 membrane synthesis, maturation, and budding. After the end of productive L2 infection and establishment of lysogeny, osmotic fragility returns to that of uninfected cells.

L2 has a noncytocidal productive infection cycle which is followed by a lysogenic cycle in most if not all L2-infected cells. The regulatory mechanism for a decision between productive and lysogenic cycles therefore must be different for L2 and the lambdoid phages. For the latter, the decision is between a lytic and a lysogenic cycle, which are mutually exclusive. Similarly, phage *ai* has a cytocidal productive infection cycle (above), so regulation of lysogeny requires a choice between either productive infection or lysogeny. The L2 regulatory mechanism cannot be a simple temporal switch from a productive infection cycle to a lysogenic one because L2 integration takes place during the rise period of the productive infection cycle, which continues for several hours after prophage formation. The nature of the L2 regulatory mechanism is not known.

Insertion Mutants and Miniphages

Two spontaneous L2 insertion mutants have been isolated from wild-type L2 stocks. Each has a genome of about 15 kbp, consisting of the 12.0 kbp L2 genome plus a 3.3 kbp insertion. The 3.3 kbp fragment is the result of transposition of two noncontiguous regions of the L2 genome, and the two L2 insertion mutants differ only in the location of the 3.3 kbp insert. During serial passage of each L2 insertion mutant, miniphage DNAs are formed. These are superhelical circular molecules of the 3.3 kbp insertion DNA and concatemers of the 3.3 kbp DNA. These miniphage DNAs can be packaged during L2 infection and produce particles that are defective and, perhaps, interfering.

Unclassified Phages

There are a number of reports of uncharacterized mycoplasma phages. Most of these have been reviewed previously (23, 24) and will not be discussed in detail here because no new information is available. Examples include phage M1 of *Acholeplasma modicum* (8), phage O1 of *A. oculi* (30), phage Br1 of *Mycoplasma bovirhinis* (17), and phage Hr1 of *Mycoplasma hyorhinis* (18). Many of the unclassified mycoplasma phages are thought to have DNA genomes based on their tailed-phage morphology, but in most cases no data are available on the nucleic acid content. In addition to these unclassified viruses, there are a numerous reports of virus-like particles in electron micrographs of mycoplasma cultures (27). Also, there are anecdotal stories of extrachromosomal elements and a few reports of spontaneous plaques on lawns of some mycoplasma species. These observations suggest that mycoplasma phages may be more common than the sporadic reports of their isolation might indicate. Of the unclassified mycoplasma phages, significant data are only available for *Mycoplasma* phage MAV1 and the *Acholeplasma* phage L172.

Mycoplasma Phage MAV1

History and Importance

Phage MAV1 infects *M. arthritidis*, a mycoplasma that causes chronic arthritis in murine animals (31, 40). Initially, a 16 kbp extrachromosomal dsDNA element was identified in *M. arthritidis* (strains 158p10, 14124, and 14152) that was suspected of being of phage origin because of its linear restriction map (49). Subsequently, *Mycoplasma* phage MAV1 was plaque-isolated from one of these *M. arthritidis* strains (158p10) using *M. arthritidis* strain 158 as the indicator strain. As a lysogenic phage, MAV1 is uniquely important for studies of *Mycoplasma* pathogenicity because MAV1 lysogens are more virulent than nonlysogens (45, 51) (see chapter 47 for a general overview of the phage impact on bacterial pathogenicity).

Virion and Macromolecules

The morphology of MAV1 is not known and the taxonomy of *Mycoplasma* phage MAV1 has therefore not been determined. MAV1 is resistant to treatment with DNase I, RNase A, chloroform, the nonionic detergent Triton X-100, trypsin, and proteinase K (49). Although many phages are relatively resistant to proteases, the resistance of a phage to a general protease, such as proteinase K, is unusual. However, *Spiroplasma* phage SpV4 (reviewed above) is also reported to be proteinase K resistant (33). Protease resistance may have evolved to adapt these phages to their host cells. Mycoplasmas have small genomes and lack many biosynthetic pathways, particularly for amino acids, purines, and pyrimidines. Hence, they must acquire these compounds from the environment by the production and secretion of scavenging proteases and nucleases. We postulate, therefore, that MAV1 has evolved resistance to a general protease secreted by *M. arthritidis* that has not been identified yet.

MAV1 has been found to have a linear dsDNA genome of 15,644 bp with 29.0% G-C content (50). The degradation of MAV1 DNA by λ exonuclease indicates that the 5′-terminus of the DNA strands is exposed and not attached to terminal protein (49). With one exception, all MAV1 ORFs are on the same DNA strand (designated the (+) strand). Amino acid sequences of most of the predicted MAV1 proteins have no significant matches with proteins in sequence databases. Exceptions are the MAV1 ORFs that are predicted to encode a DNA replicase, an integrase, and a cytosine-specific restriction-modification system. Another ORF (the *imm* gene) would encode a protein with motifs characteristic of phage repressor proteins and may be the immunity protein that renders MAV1 lysogens resistant to superinfection. The lone ORF on the (−) strand is predicted to encode a lipoprotein designated Vir.

An interesting aspect of the MAV1 genome sequence, like that of *Acholeplasma* phage L3 (above), is the complete absence of the sequence GATC, which would indicate that phage MAV1 has evolved to avoid GATC-specific restriction-modification systems. However, cell chromosomal DNA of the strains of *M. arthritidis* that have been examined is readily degraded by the GATC-specific restriction enzymes, *Mbo*I and *Sau*3AI (49). It has been suggested that the MAV1 hosts that have been studied lack GATC-specific restriction systems but that other hosts that have not yet been identified may have such systems. Alternatively, rather than an unknown host having a GATC-specific restriction system, the cytosine-specific restriction system which is predicted to be encoded by the MAV1 genome may recognize GATC. MAV1 lysogens have unmodified GATC sequences that are cleaved by *Mbo*I and *Sau*3AI, but this would be expected because a hypothetical phage-encoded GATC-specific restriction system would most likely be expressed only during productive infection and not in lysogens. In support of this latter possibility, reverse transcription polymerase chain reaction (RT-PCR) analysis of total RNA isolated from MAV1 lysogens detected mRNA transcripts corresponding to the *imm* and *vir* genes but not to other MAV1 genes, such as those encoding the restriction system (50).

Growth and Replication

MAV1 growth curves have not been done. One of the difficulties with studying MAV1 is that the phage has not been reproducibly propagated in broth. When *M. arthritidis* cells that are susceptible to MAV1 are mixed with the phage in broth, essentially 100% of the cells become lysogens with the production of few phage particles, even when the multiplicity of infection is low. In contrast to *M. arthritidis* in broth media, MAV1 is readily propagated as plaques on lawns in agar. Stocks of the phage can be prepared with titers of 10^8 PFU/ml by scraping top agar overlays from lawns of host cells that have undergone confluent lysis (49).

Lysogeny

Based on Southern analysis using probes specific for MAV1 DNA, about half of the *M. arthritidis* strains that have been examined are MAV1 lysogens (51). Strains containing MAV1 DNA sequences were resistant to infection with MAV1. Most but not all of the nonlysogenic strains were found to be MAV1 hosts. In some cases, *M. arthritidis* strains that had been maintained in different laboratories differed in whether they were MAV1 lysogens. We suspect that *M. arthritidis* strains have on occasion been unknowingly infected with MAV1, generating strain differences with regard to lysogeny among laboratories.

During lysogeny, MAV1 DNA inserts into the genome of *M. arthritidis* at the site TATTTTT (49). This 7 bp sequence is common in the AT-rich genome of *M. arthritidis* and is present at 42 sites in the MAV1 genome. MAV1 DNA can therefore integrate at numerous sites in the host chromosome during the establishment of lysogeny.

The mechanism by which MAV1 DNA integrates into the *M. arthritidis* chromosome is not known. MAV1 DNA probably forms a circular integration intermediate, but this has not been proven. PCR experiments using primers oriented outward from the ends of the linear MAV1 genome yield products consistent with some template molecules being circular, but the products also could arise from template molecules that are head-to-tail dimers (49). Restriction mapping data show that the termini of linear MAV1 genomic DNA isolated from virions are the same as the termini of MAV1 prophages in lysogens. The ends of the MAV1 genome when circularized form the 7 bp TATTTTT sequence. Thus, it appears that MAV1 DNA inserts into the chromosome by a site-specific recombination mechanisms at the TATTTTT sequence.

Some features of MAV1 DNA integration resemble the insertion of transposon Tn*916* during transposition. Sequences of the extreme ends of MAV1 prophages are not identical in all lysogens (49). Single nucleotide differences at the MAV1 DNA ends could be interpreted as evidence of an integration mechanism involving a staggered nucleotide cleavage. A circular intermediate during Tn*916* transposition also has the sequence TATTTTT at the recombination site, and a variable sequence (sometimes referred to as a coupling sequence) at the ends of the transposon varies as a result of the staggered strand exchange mechanism (5).

Impact on Host Virulence

The association between MAV1 and *M. arthritidis* virulence is clear, but the underlying mechanism requires further study. Preliminary data indicate that the mycoplasma load, assayed as CFU, is greater in animals infected with MAV1 lysogens than in animals infected with nonlysogens (A.-H. T. Tu and K. Dybvig, unpublished data). The difference in mycoplasma load is apparent as little as 3 hours post-inoculation. Therefore, MAV1 lysogens may resist innate host defenses more effectively than nonlysogens.

Because MAV1 lysogens are always more virulent than nonlysogens (51), regardless of the particular TATTTTT site into which MAV1 DNA has integrated, it has been proposed that MAV1 encodes a virulence factor (50). The MAV1 Vir protein is one of only a few examples of lipoproteins encoded by phage genomes. MAV1 virions are resistant to treatment with nonionic detergents and chloroform and, therefore, lack lipid. Thus, it is assumed that Vir is anchored in the cell membrane of MAV1-lysogenized *M. arthritidis* where it may interact with *M. arthritidis*-infected organisms. Vir may function similarly to the lipoprotein of bacteriophage λ, Bor, which is involved in serum resistance (1).

Acholeplasma *Phage L172*

History and Importance

Acholeplasma phage L172 was isolated from washes of *A. laidlawii* lawns and, based on morphological observations, was believed to be related to *Acholeplasma* phage L2, an enveloped, quasi-spherical phage containing circular dsDNA (above). This led to erroneous interpretations of subsequent biochemical and biophysical studies "confirming" L172 to be a circular dsDNA phage. Later investigators, noting that L172 DNA was sensitive to various nucleases not expected to hydrolyze for dsDNA, showed that L172 contains circular ssDNA; the original studies apparently had been done using virion ssDNA contaminated with RF dsDNA.

Virion and Macromolecules

L172 virions are enveloped quasi-spherical particles, 60–80 nm in diameter. About 10–12% of L172 dry weight is lipid, consisting of phospholipids, glycolipids, and phosphoglycolipids. L172 membrane lipid and fatty acid compositions are similar to those of the *A. laidlawii* membrane. The virion membrane is 6.5–8.0 nm, but no internal structure has been observed. Due to the limited data available, the taxonomy of *Acholeplasma* phage L172 has not been established, although L172 clearly represents a new phage family: enveloped, quasi-spherical phages containing ssDNA (23). L172 is inactivated by the nonionic detergents octal glucoside and Triton X-100, and by ether and it is also sensitive to heat inactivation.

The L172 genome is 14 kb ssDNA with a G-C content of 29.4%, significantly less than the 31.8% G-C content of *A. laidlawii* host cells and the 32.0% G-C content of *Acholeplasma* phage L2. L172 DNA is sensitive to S1 nuclease and resistant to exonuclease VII. L172 virions, analyzed by SDS-PAGE, contain seven proteins of 71, 68, 53, 42, 40, 18, and 15 kDa.

Growth and Replication

The limited data available indicate that replication of L172 ssDNA in *A. laidlawii* host cells is similar to that of L51 ssDNA: L172 RF dsDNA has been isolated from infected cells and L172 cannot infect *A. laidlawii* REP$^-$ cells, which are mutants that cannot be infected by ssDNA *Acholeplasma* phages but can be infected by dsDNA *Acholeplasma* phages.

Host Restriction and Modification

Many restriction-modification systems have been reported in mycoplasmas. Previously reviewed restriction systems include the CAATTG-specific enzyme of *Mycoplasma fermentans*, a GATC-specific system in *A. laidlawii*, another *A. laidlawii* system that restricts DNA containing 5-methylcytosine regardless of sequence context, a GCGC-specific system in *S. citri*, and a GCNGC-specific system in *Ureaplasma urealyticum* (13, 26). Some mycoplasmas have potential restriction-modification systems based on analysis of complete genome sequences, but phenotypic data are lacking (15, 20).

Mar*I Restriction System*

Many *M. arthritidis* strains have a restriction system that recognizes AGCT. This system is designated *Mar*I in the REBASE database (http://rebase.neb.com/rebase/rebase.html) (48). *Mar*I is an isoschizomer of the *Alu*I restriction enzyme, and *M. arthritidis* genomic DNA is resistant to cleavage by *Alu*I. Initial attempts to genetically transform *M. arthritidis* strains that possess *Mar*I were unsuccessful. Transformation was achieved only when the transforming DNA had been modified in vitro with *Alu*I DNA methyltransferase (MTase) (48). Thus, *Mar*I is a significant barrier to gene transfer in *M. arthritidis*. Genomic DNA of all *Mycoplasma* phage MAV1 host strains that have been examined is resistant to *Alu*I digestion, indicating *Mar*I is widespread in *M. arthritidis*. Phage MAV1 DNA has four AGCT sites and would be restricted if it were not modified by the *Mar*I MTase. However, because the known MAV1 hosts all possess *Mar*I, all MAV1 phage stocks are modified at AGCT sites and, therefore, not restricted during infection.

Phase-Variable Restriction Systems

M. pulmonis hsd Loci

M. pulmonis has complex DNA inversion systems that encode type I restriction and modification enzymes. Type I enzymes contain two HsdS subunits that determine the DNA recognition sequence: HsdM subunits that contain the catalytic domain for the MTase reaction and an HsdR subunit with the catalytic domain for DNA cleavage. Although both initial binding of the enzyme to duplex DNA and the MTase reaction are site-specific, the DNA cleavage reaction occurs at essentially random sites due to the enzyme's ATP-dependent DNA translocation activity. Each *M. pulmonis hsd* locus has two *hsdS* genes, one *hsdR* gene, and one *hsdM* gene (14). The two *hsdS* genes flank *hsdR* and *hsdM* and are in an inverted orientation to one another. DNA inversions occur at high frequency between the two *hsdS* genes (12, 14, 43). The coding region of each *hsdS* gene has two or three recombination sites for DNA inversions. Thus, the primary amino acid sequences of the HsdS proteins vary according to the particular DNA inversion that occurs. These changes in HsdS sequence alter the specificity of the DNA recognition sequence of the type I enzyme, as assessed by the plaquing efficiency of *Mycoplasma* phage P1 (12). Therefore, DNA inversions between *hsdS* sequences generate a family of *M. pulmonis* restriction and modification enzymes with differing specificities.

The *hsd* DNA inversions not only change the specificities of HsdS subunits but also result in phase-variable expression of restriction and modification activity (12, 14). One of the *hsdS* genes along with *hsdR* and *hsdM* are transcribed as an operon (43). There is a single *hsd* promoter at one end of the *hsd* locus that drives transcription of the *hsdS*, *hsdR*, and *hsdM* operon or from the second *hsdS* gene, depending on the orientation of the locus in the chromosome. Thus, one of the *hsdS* genes is always transcribed and expression of *hsdR* and *hsdM* is phase-variable. In the absence of *hsdR* and *hsdM* transcription, cells lack restriction and modification activity. *M. pulmonis* strain KD735 has two *hsd* loci: *hsd1* and *hsd2* (43). If either locus is oriented such that *hsdR* and *hsdM* are transcribed, two restriction and modification enzymes are produced (12). One enzyme has the HsdS subunit encoded by the transcribed *hsd1* gene, and the other enzyme's HsdS subunit is encoded by the transcribed *hsd2* gene. Cells that lack restriction and modification activity only arise when both loci are oriented such that *hsdR* and *hsdM* are not transcribed.

The complete genome sequence of *M. pulmonis* strain CT reveals a third *hsd* locus, *hsd3*, not identified in strain KD735 (6). The *hsdR* and *hsdM* genes of *hsd3* are most likely defective due to frameshift mutations, but the *hsdS* genes appear functional. It is possible that three restriction and modification enzymes would be produced in the CT strain when *hsdR* and *hsdM* are transcribed because an *hsdS* gene could be transcribed from each of the three *hsd* loci.

The *M. pulmonis* genome sequence (http://genolist.pasteur.fr/MypuList/) has a gene (MYPU 5310) near the *vsa* locus that is predicted to encode a site-specific DNA recombinase. The *vsa* locus encodes a family of surface lipoproteins (V-1 antigens) that vary due to site-specific DNA inversions within the *vsa* locus (3, 39, 41). Recently, an *M. pulmonis* mutant has been isolated from a transposon library in which the recombinase gene is disrupted (42). The mutant fails to undergo DNA inversions at both *vsa* and *hsd* loci. *E. coli* cells containing the cloned recombinase gene, but not *E. coli* that lack the recombinase gene, can undergo site-specific DNA inversions between both *vsa* and *hsd* sequences (42). Thus, it appears that a single site-specific DNA recombinase catalyzes both *vsa* and *hsd* inversions, indicating that antigenic variation and restriction enzyme variation is linked. Such linkage may explain, at least in part, why a greater percentage of *M. pulmonis* organisms isolated from the respiratory tract of infected animals have active restriction systems than do organisms that have been maintained in laboratory media (19).

Other Putative Phase-Variable Enzymes in *M. pulmonis*

Several putative *M. pulmonis* genes, or partial genes, are predicted to encode type III DNA MTase enzymes. These are designated in the genome sequence as MYPU 3950, 3960, 3970, 3980, and 4800. Other genes that are predicted to encode cytosine-specific MTase enzymes are MYPU 0430, 0440, 1850, and 1860. MYPU 4720 and 6880 are predicted to encode adenine-specific MTase enzymes.

Many of these genes have tandem dinucleotide repeats within the coding region. The number of tandem repeats would likely vary as a result of slipped-strand mispairing during DNA replication (22). MYPU 1850, 3960, 3970, and 3980 have 9, 8, 12 and 7 tandem AG repeats, respectively; MYPU 4800 has 13 tandem GA repeats; and MYPU 0440 has 14 tandem CA repeats. For some of these genes, the ORF would be extended if the number of tandem repeats increases or decreases by one. MYPU 3960 and 3970, for example, would be extended 231 and 237 nucleotides, respectively, if the number of tandem AG repeats increases by one. If the 9 AG repeats in MYPU 1850 increased to 10, then the resulting frameshift would merge MYPU 1850 (predicted gene product of 82 amino acids) and 1860 (predicted gene product of 264 amino acids) into a single ORF encoding a functional MTase of 432 amino acids. Similarly, the frameshift resulting from a change in the number of CA repeats in MYPU 0440 from 14 to 15 would merge MYPU 0440 and MYPU 0430 into a single gene that would most likely encode a functional MTase of 406 amino acids. Accordingly, it is predicted that *M. pulmonis* possesses several MTase enzymes, some of which would be phase-variable due to slipped-strand mispairing.

The predicted proteins that would result from merging MYPU 1850 and 1860 and from merging MYPU 1430 and 1440 by slipped-strand mispairing share significant amino acid sequence similarity with one another and with numerous cytosine-specific MTase enzymes from other organisms including some mycoplasmas. Examples are putative enzymes from *M. mycoides* subsp. *capri* (GenBank Accession No. AF072715) and *U. urealyticum* (15). Another example is the *Sss*I MTase from *Spiroplasma sp.* strain MQ1 that methylates cytosine in the sequence CG (34).

A Possible *Mycoplasma hominis* Phase-Variable Mtase

The cytosine-specific MTase that is predicted to be encoded by the *Mycoplasma* phage MAV1 genome is similar to the predicted gene product from a partial ORF found upstream of the *M. hominis dnaK* gene (GenBank Accession No. AJ132792) (K. Dybvig, unpublished data). In addition to predicted amino acid sequence similarity, the MAV1 and *M. hominis* genes share extensive nucleotide similarity. *M. hominis* and *M. arthritidis*, the phage MAV1 host, are phylogenetically related and it is possible that MAV1-related phages are responsible for the spread of restriction and modification genes among *Mycoplasma* species through transduction. The *M. hominis* MTase ORF lacks a convincing ribosome binding site and a translation start site. However, near the 5′-end of the coding region are 10 tandem copies of the dinucleotide AG. If the number of AG repeats changes to 11 by slipped-strand mispairing, a ribosome binding site and translation start site would emerge. Thus, production of the *M. hominis* MTase may be phase-variable.

Functional Significance of Phase-Variable Restriction and Modification Enzymes

The purpose of phase-variable restriction systems is unknown. In *M. pulmonis* laboratory-adapted strains, nearly all cells in a culture have *hsd* loci oriented such that *hsdR* and *hsdM* are not transcribed. Thus, the majority of cells in culture are readily susceptible to phage P1. However, the primary function of the *hsd* loci may not be to protect cells from phage infection. This is especially true for the various other phase-variable MTase enzymes that *M. pulmonis* is predicted to produce. No cognate restriction endonucleases for these MTase enzymes were identified from the complete genome sequence. Finally, it is not known how genomic DNA escapes endonucleolytic attack when phase-variable restriction activity is induced by *hsd* inversion. Possible answers to some of these questions have been proposed but are not altogether satisfying (9, 19, 21, 43).

Conclusions

The unusual phenotypes of mycoplasma phages may reflect the rapid rate of evolution that has characterized the degenerate evolution of mycoplasmas from Gram-positive bacteria (25). This may have allowed mycoplasma cells and phages to explore evolutionary alternatives not accessible to other biological systems. In some cases mycoplasma phages seem to have evolved new infection strategies, like the distinctive type of temperate infection of *Acholeplasma* phage L2 (productive infection and lysogeny in the same infected cell). In other cases mycoplasma phages seem to be phages that have adapted from growth in eubacteria to growth in mycoplasmas, like the short-tailed phages that infect *Acholeplasma*, *Spiroplasma*, and *Mycoplasma* species. For these phages, there remain unanswered questions on the role of phage tails in infections involving wall-less mycoplasma host cells and the mechanism of nonlytic release of progeny tailed phages from infected cells.

An unexpected similarity has been reported in the capsid protein structure of the small icosahedral ssDNA phages that infect *Spiroplasma*, *Bdellovibrio*, and *Chlamydia* (4). *Spiroplasma* are evolutionarily degenerate Gram-positive bacteria without cell walls. However, *Bdellovibrio* and *Chlamydia* are Gram-negative bacteria, each with distinctive morphological, growth, and biochemical characteristics, and phylogenetically distant from each other and from *Spiroplasma*. If the similarity in the capsid proteins of the *Microviridae* phages that infect these three genera is an example of convergent evolution, this would imply similar selective pressures presented by the different types of cell surfaces of these phylogenetically very distant host genera.

There are now a couple of examples of mycoplasma phages affecting host pathogenicity. The increased virulence observed when *M. arthritidis* is a phage MAV1 lysogen provides a new approach for investigating poorly understood mycoplasma diseases. Also, preliminary data on the effect of infections with heterogeneous SRO phages in eliminating SRO spiroplasmas causing abnormal sex ratios in *Drosophila* may be an example of phage therapy (phage therapy is reviewed in chapter 48).

The phylogeny of the three mycoplasma genera that are known phage hosts shows that *Acholeplasma* arose from the *Streptococcus* phylogenetic branch, *Spiroplasma* arose from the *Acholeplasma* phylogenetic branch about the time of the first land plants and insects, and *Mycoplasma* arose from the *Spiroplasma* phylogenetic branch about the time of the first mammals. Questions about the origin and evolution of mycoplasma phages can be phrased in terms of this host phylogeny. Did mycoplasma phages arise from phages of Gram-positive eubacteria and coevolve with *Acholeplasma* species as they diverged from the *Streptococcus* phylogenetic branch, then diverge later and coevolve with *Spiroplasma* species, and finally diverge again and coevolve with *Mycoplasma* species? In this model, did mycoplasma phages play a role in host adaptation to growth in new ecological niches during evolution of the biosphere? The requirement that *Spiroplasma* and *Mycoplasma* phages use UGA as a tryptophan codon argues against a recent origin of these phages and suggests they must have arisen along with *Spiroplasma* and coevolved the necessary codon changes with these host cells.

References

1. Barondess, J. J., and J. Beckwith. 1995. *bor* gene of phage λ, involved in serum resistance, encodes a widely conserved outer membrane lipoprotein. J. Bacteriol. 177:1247–1253.
2. Bébéar, C.-M., P. Aullo, J.-M. Bové, and J. Renaudin. 1996. *Spiroplasma citri* virus SpV1: characterization of viral sequences present in the spiroplasmal host chromosome. Curr. Microbiol. 32:134–140.
3. Bhugra, B., L. L. Voelker, N. Zou, H. Yu, and K. Dybvig. 1995. Mechanism of antigenic variation in *Mycoplasma pulmonis*: interwoven, site-specific DNA inversions. Mol. Microbiol. 18:703–714.
4. Brentlinger, K. L., S. Hafenstein, C. R. Novak, B. A. Fane, R. Borgon, R. McKenna, and M. Agbandje-McKenna. 2002. *Microviridae*, a family divided: isolation, characterization, and genome sequence of ϕMH2K, a bacteriophage of the obligate intracellular parasite bacterium *Bdellovibrio bacteriovorus*. J. Bacteriol. 184:1089–1094.
5. Caparon, M. G., and J. R. Scott. 1989. Excision and insertion of the conjugative transposon T*n916* involves a novel recombination mechanism. Cell 59:1027–1034.
6. Chambaud, I., R. Heilig, S. Ferris, V. Barbe, D. Samson, F. Galisson, I. Moszer, K. Dybvig, H. Wróblewski, A. Viari, E. P. C. Rocha, and A. Blanchard. 2001. The complete genome of the murine respiratory pathogen *Mycoplasma pulmonis*. Nucleic Acids Res. 29:2145–2153.
7. Chipman, P. R., M. Agbandje-McKenna, J. Renaudin, T. S. Baker, and R. McKenna. 1998. Structural analysis of the spiroplasma virus SpV4: implications for evolutionary variation to obtain host diversity among the *Microviridae*. Structure 6:135–145.
8. Congdon, A. L., E. S. Boatman, and G. E. Kenny. 1979. Mycoplasmatales virus MV-M1: discovery in *Acholeplasma modicum* and preliminary characterization. Curr. Microbiol. 3:111–115.
9. Dybvig, K. 1993. DNA rearrangements and phenotypic switching in prokaryotes. Mol. Microbiol. 10:465–471.
10. Dybvig, K., J. Alderete, and G. H. Cassell. 1988. Adsorption of *Mycoplasma pulmonis* virus P1 to host cells. J. Bacteriol. 170:4373–4375.

11. Dybvig, K., A. Liss, J. Alderete, R. M. Cole, and G. H. Cassell. 1987. Isolation of a virus from *Mycoplasma pulmonis*. Isr. J. Med. Sci. 23:418–422.
12. Dybvig, K., R. Sitaraman, and C. T. French. 1998. A family of phase-variable restriction enzymes with differing specificities generated by high-frequency gene rearrangements. Proc. Natl. Acad. Sci. USA 95:13923–13928.
13. Dybvig, K., and L. L. Voelker. 1996. Molecular biology of mycoplasmas. Annu. Rev. Microbiol. 50:25–57.
14. Dybvig, K., and H. Yu. 1994. Regulation of a restriction and modification system via DNA inversion in *Mycoplasma pulmonis*. Mol. Microbiol. 12:547–560.
15. Glass, J. I., E. J. Lefkowitz, J. S. Glass, C. R. Heiner, E. Y. Chen, and G. H. Cassell. 2000. The complete sequence of the mucosal pathogen *Ureaplasma urealyticum*. Nature 407:757–762.
16. Gourlay, R. N. 1970. Isolation of a virus infecting a strain of *Mycoplasma laidlawii*. Nature 25:1165.
17. Gourlay, R. N., S. G. Wyld, and D. J. Garwes. 1983. Some properties of mycoplasma virus Br1. Arch. Virol. 75:1–15.
18. Gourlay, R. N., S. G. Wyld, and M. E. Poulton. 1983. Some characteristics of mycoplasma virus Hr1, isolated from and infecting *Mycoplasma hyorhinis*. Arch. Virol. 77:81–85.
19. Gumulak–Smith, J., A. Teachman, A.-H. T. Tu, J. W. Simecka, J. R. Lindsey, and K. Dybvig. 2001. Selection pressure for variations in Vsa surface proteins and restriction enzyme systems of *Mycoplasma pulmonis* during animal infection. Mol. Microbiol. 40:1037–1044.
20. Himmelreich, R., H. Hilbert, H. Plagens, E. Pirkl, B.-C. Li, and R. Herrmann. 1996. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. Nucleic. Acids Res. 24:4420–4449.
21. Janscak, P., D. T. F. Dyden, and K. Firman. 1998. Analysis of the subunit assembly of the type IC restriction-modification enzyme *Eco*R124I. Nucleic Acids Res. 26:4439–4445.
22. Levinson, G., and G. A. Gutman. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol. 4:203–221.
23. Maniloff, J. 1988. Mycoplasma viruses. Crit. Rev. Microbiol. 15:339–389.
24. Maniloff, J. 1992. Mycoplasma viruses, pp. 41–59. *In* J. Maniloff, R. N. McElhaney, L. R. inch, and J. B. Baseman (eds.), Mycoplasmas: Molecular Biology and Pathogenesis. American Society for Microbiology, Washington, D.C.
25. Maniloff, J. 2002. Phylogeny and evolution, pp. 31–43. *In* S. Razin, and R. Herrmann (eds.) Molecular Biology and Pathogenicity of Mycoplasmas. Kluwer Academic/Plenum Publishers, New York.
26. Maniloff, J., K. Dybvig, and T. L. Sladek. 1992. Mycoplasma DNA restriction and modification, pp. 325–330. *In* J. Maniloff, R. N. McElhaney, L. R. Finch, and J. B. Baseman (eds.) Mycoplasmas: Molecular Biology and Pathogenesis. American Society for Microbiology, Washington, D.C.
27. Maniloff, J., K. Haberer, R. N. Gourlay, J. Das, and R. Cole. 1982. Mycoplasma viruses. Intervirology 18:177–188.
28. Maniloff, J., G. J. Kampo, and C. C. Dascher. 1994. Sequence analysis of a unique phage: mycoplasma virus L2. Gene 141:1–8.
29. Melcher, U., Y. Sha, F. Ye, and J. Fletcher. 1999. Mechanisms of spiroplasma genome variation associated with SpV1-like viral DNA inferred from sequence comparisons. Microbial Comp. Genomics 4:29–46.
30. Ogawa, H. I., and M. Nakamura. 1985. Characterization of a mycoplasma virus (MV-O1) derived from and infecting *Acholeplasma oculi*. J. Gen. Microbiol. 131:3117–3126.
31. Razin, S., D. Yogev, and Y. Naot. 1998. Molecular biology and pathogenicity of mycoplasmas. Microbiol. Mol. Biol. Rev. 62:1094–1156.
32. Renaudin, J., and J. M. Bové. 1994. SpV1 and SpV4, spiroplasma viruses with circular, single-stranded DNA gemomes, and their contribution to the molecular biology of spiroplasmas. Adv. Virus Res. 44:429–463.
33. Renaudin, J., M. C. Pasarel, M. Garnier, P. Carle, and J. M. Bové. 1984. Characterization of spiroplasma group 4 (SV4). Isr. J. Med. Sci. 20:797–799.
34. Renbaum, P., D. Abrahamove, A. Fainsod, G. G. Wilson, S. Rottem, and A. Razin. 1990. Cloning, characterization, and expression in *Escherichia coli* of the gene coding for the CpG DNA methylase from *Spiroplasma sp.* strain MQ1 (M$SssI). Nucleic Acids Res. 18:1145–1152.
35. Salas, M. 1991. Protein-priming of DNA replication. Annu. Rev. Biochem. 60:39–71.
36. Salas, M., and F. Rojo. 1993. Replication and transcription of bacteriophage ϕ29 DNA, pp. 843–857. *In* A. L. Sonenshein, J. A. Hoch, and R. Losick (eds.) *Bacillus subtilis* and other Gram-positive Bacteria. American Society for Microbiology, Washington, D.C.
37. Sha, Y., U. Melcher, R. E. Davis, and J. Fletcher. 1995. Resistance of *Spiroplasma citri* lines to the virus SVTS2 is associated with integration of viral DNA sequences into host chromosomal and extrachromosomal DNA. Appl. Environ. Microbiol. 61:3950–3959.
38. Sha, Y., U. Melcher, R. E. Davis, and J. Fletcher. 2000. Common elements of spiroplasma plectroviruses revealed by nucleotide sequence of SVTS2. Virus Genes 20:47–56.
39. Shen, X., H. Yu, J. Gumulak, C.T. French, N. Zou, and K. Dybvig. 2000. Gene rearrangements in the *vsa* locus of *Mycoplasma pulmonis*. J. Bacteriol. 182:2900–2908.
40. Simecka, J. W., J. K. Davis, M. K. Davidson, S. E. Ross, C. T. K.-H. Stadtlander, and G. H. Cassell. 1992. Mycoplasma diseases of animals, pp. 391–415. *In* J. Maniloff, R. N. McElhaney, L. R. Finch, and J. B. Baseman (eds) Mycoplasmas: molecular biology and Pathogenesis. American Society for Microbiology, Washington, D.C.
41. Simmons, W. L., C. Zuhua, J. I. Glass, J. W. Simecka, G. H. Cassell, and H. L. Watson. 1996. Sequence analysis of the chromosomal region around and within the V-1-encoding gene of *Mycoplasma pulmonis*: evidence for DNA inversion as a mechanism for V-1 variation. Infect. Immun. 64:472–479.
42. Sitaraman, R., A. M. Denison, and K. Dybvig. 2002. A unique, bifunctional site-specific DNA recombinase from *Mycoplasma pulmonis*. Mol. Microbiol. 46:1033–1040.
43. Sitaraman, R., and K. Dybvig. 1997. The *hsd* loci of *Mycoplasma pulmonis:* organization, rearrangements and expression of genes. Mol. Microbiol. 26:109–120.

44. Smith, M. C. M., N. Burns, J. R. Sayers, J. A. Sorrell, S. R Casjens, and R. W. Hendrix. 1998. Bacteriophage collagen. Science 279:1834.
45. Tu, A.-H. T., J. R. Lindsey, T. R. Schoeb, A. Elgavish, H. Yu, and K. Dybvig. 2002. Bacteriophage MAV1 is a mycoplasmal virulence factor for the development of arthritis in both rats and mice. J. Infect. Dis. 186:432–435.
46. Tu, A.-H. T., L. L. Voelker, X. Shen, and K. Dybvig. 2001. Complete nucleotide sequence of the mycoplasma virus P1 genome. Plasmid 45:122–126.
47. Van Regenmortel, M. H. V., C. M. Fauquet, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, J. Maniloff, M. A. Mayo, D. J. McGeoch, C. R. Pringle, and R. B. Wickner (eds.). 2000. Virus Taxonomy: Seventh Report of the International Committee on Taxonomy of Viruses. Academic Press, New York.
48. Voelker, L. L., and K. Dybvig. 1996. Gene transfer in *Mycoplasma arthritidis*: transformation, conjugal transfer of Tn*916*, and evidence for a restriction system recognizing AGCT. J. Bacteriol. 178:6078–6081.
49. Voelker, L. L., and K. Dybvig. 1998. Characterization of the lysogenic bacteriophage AV1 from *Mycoplasma arthritidis*. J. Bacteriol. 180:5928–5931.
50. Voelker, L. L., and K. Dybvig. 1999. Sequence analysis of the *Mycoplasma arthritidis* bacteriophage MAV1 genome identifies the putative virulence factor. Gene 233:101–107.
51. Voelker, L. L., K. E. Weaver, L. J. Ehle, and L. R. Washburn. 1995. Association of lysogenic bacteriophage MAV1 with virulence of *Mycoplasma arthritidis*. Infect. Immun. 63:4016–4023.
52. Zou, N., K. Park, and K. Dybvig. 1995. Mycoplasma virus P1 has a linear, double-stranded DNA genome with inverted terminal repeats. Plasmid 33:41–49.

41

Lactobacillus Phages

HARALD BRÜSSOW
JUAN E. SUÁREZ

The studies conducted with *Lactobacillus* bacteriophages reflect the economic and medical importance of their hosts. Due to the variety of food fermentation lactobacilli-based processes that can become disrupted by phage development, and the diversity of mucosal surfaces colonized by lactobacilli, phage research has not concentrated on a single *Lactobacillus* species or phage. *Lactobacillus* phage research also is frequently directed by practical needs. For example, the industrially relevant properties of lactic acid bacteria are often plasmid-encoded. These plasmids are intrinsically unstable, which frequently leads to strain degeneration. The DNA-integration mechanism used by temperate *Lactobacillus* phages therefore has been studied to develop tools for chromosomal stabilization of economically important traits (57, 73). In addition, many fermented products undergo a ripening period that may last several months. Early lysis of the starter bacteria might accelerate ripening through release of the intracellular enzymes into the food matrix. To this aim, the expression of cloned phage lysis cassettes, controlled by inducible promoters, has been studied (19). Furthermore, the need of phage-insensitive *Lactobacillus*-based fermentation starters has promoted the generation of strains harbouring a *cI*-like repressor gene or a phage replication origin (3, 59).

Phage Types

All types of tailed phages (*Caudovirales*) have been isolated from lactobacilli. The phages with long, noncontractile tails (*Siphoviridae*) come in two classes: phages with isometric heads represent the majority of the isolates (figure 41-1A) and are relatively well documented by complete genome sequences (1, 46, 58, 66). In contrast, phages with prolate heads and peculiar knob-like appendages along the tail (e.g., the group c *Lactobacillus delbrueckii* phage JCL1032; 26) have not yet been characterized in any molecular detail. Molecular data on *Lactobacillus* phages with contractile tails (*Myoviridae*) and more complicated tail appendages (figure 41-1B) are starting to accumulate. Examples of *Podoviridae Lactobacillus* phages have not yet been investigated. See chapter 2 for a general overview of phage classification.

Myoviridae

Myoviridae have been isolated from *Lactobacillus casei* (37) and *Lactobacillus plantarum*. They come in two genome size classes: the smallest are 40–55 kb in size and the biggest about 130 kb. Among the latter, *L. plantarum* phage ΦLP65 was isolated from a meat fermentation and sequenced (16b). The genome showed about 160 open reading frames (*orfs*). N-terminal sequencing and mass spectrometric analysis of the major structural proteins from the phage ΦLP65 virion allowed the identification of the structural module on its genome map.

The gene map of the structural module was identical to that of the myovirus A511 that infects *Listeria*, a phylogenetic relative of *Lactobacillus* (see chapter 37). The two phages shared sequence identity at the protein level (25–56% amino acid identities) but not at the DNA sequence level. The overall genetic organization of this module (terminase, portal, major head, major tail, tail tape measure, side tail fiber, lysis genes) still resembled that of Siphoviridae from the same group of bacteria (see below). DNA sequence similarity was, however, only detected for the tape measure, lysin, and a few putative transcriptional regulation genes from *Lactobacillus* siphoviruses. The regulation genes were located at both genome ends of phage ΦLP65. Downstream of the ΦLp65 lysin lies a cluster of 13 tRNA genes followed by a DNA replication module, which was identified by bioinformatic analysis. The matches included a DNA polymerase related to that of *Bacillus subtilis* phage SPO1 and endo- and exonuclease, helicase and

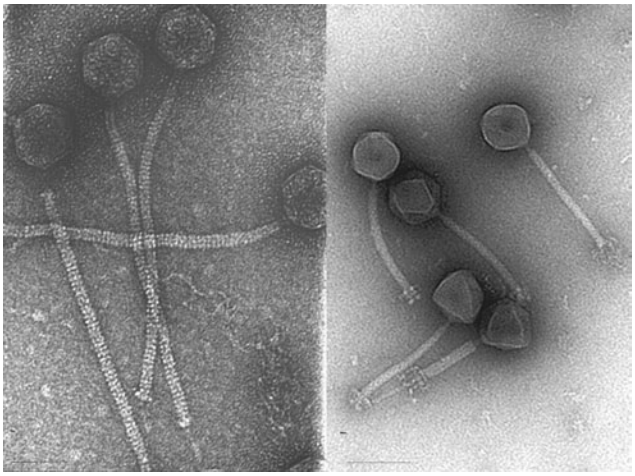


Figure 41-1 Electron micrographs of *L. casei* siphovirus A2 and *L. plantarum* myovirus ΦLP65. Head edge lengths: 60 ± 3 nm. Tail dimensions of A2:280 nm long, 12 nm wide. Tail of ΦLP65: non-contracted, 193 ± 8 nm; contracted, 115 ± 5 nm.

primase genes whose best matches were with T4-like *Myoviridae* from *Escherichia coli* (phage T4 is reviewed in chapter 18). However, more than 50% of the ORFs lacked database matches.

Siphoviridae

The isometric-head phages from lactobacilli show a familiar genome organization and belong to one of two major types of temperate phages isolated from a wide range of low G-C content Gram-positive bacteria: Sfi21-like *cos*-site *Siphoviridae* and ϕSfi11-like *pac*-site *Siphoviridae* (12). Both basic phage types showed distant relatedness with lambdoid coliphages suggesting a λ supergroup of phages. Sfi21-like phages shared the organization of the structural genes with *E. coli* phage HK97 (20, 33) while Sfi11-like phages showed weak sequence similarity to head genes from *E. coli* phage λ (22, 54) (see chapter 27 for a review of lambdoid phages and phage λ).

Cos-site *Siphoviridae*: ϕadh and A2

Two *cos*-site Sfi21-like *Lactobacillus* phages have been sequenced: *Lactobacillus gasseri* phage ϕadh (1) and *L. casei* phage A2 (30, 66). Numerous protein sequence similarities linked these two *Lactobacillus* phages and the *Streptococcus thermophilus* phage Sfi21. The similarity with streptococcal phages was especially marked for the DNA packaging and head and tail genes of phage ϕadh (21) and the DNA replication genes of phage A2 (59) (figure 41-2A). More distant relationships were detected with other Sfi21-like phages. In order of decreasing relatedness these were *Lactococcus lactis* phage BK5-T, *Staphylococcus aureus* phage PVL, *B. subtilis* phage ϕ105 and *Clostridium perfringens* phage ϕ3626. This series reflects the order of phylogenetic relatedness of the bacterial hosts, but models of coevolution of phages with their hosts have been questioned by recent data from lactococci (66). The overall genome size of the *cos*-site *Lactobacillus* phages (43 kb and about 60 *orfs*) and their modular organization is comparable to that of the *pac*-site *Lactobacillus* phages. However, a few differences are notable such as the consistent finding of the genetic linkage of the DNA packaging type and the constellation of the head morphogenesis genes.

The head gene cluster (small subunit terminase, large subunit terminase, portal protein, protease, major head protein) is characteristic for Sfi21-like *Siphoviridae* (20). These phages frequently showed proteolytic processing of the major head protein during capsid assembly. A potential maturation protease of the ClpP family was identified in the gene preceding the major head gene (29; B. Henrich, personal communication). During head maturation an N-terminal peptide of about 120 amino acids (predicting a conspicuous coiled-coil structure) is cleaved off. The released protein is supposed to serve as a scaffold protein for head morphogenesis as demonstrated in the much better investigated *E. coli* phage HK97 (33). In fact, *Lactobacillus* phage ϕadh shared not only a related gene map, but even sequence similarity with several genes from the head gene module of *Pseudomonas* phage D3 (32), suggesting an even wider distribution of the Sfi21-like siphoviruses (22).

In other respects phages ϕadh and A2 differ from the standard genome map of Sfi21-like phages. They contain many very small *orfs* around (ϕadh) or downstream (A2) of the DNA replication module, mostly without database matches. The presence of endonuclease-like genes supports the suspicion that selfish DNA elements might be found in these undefined gene clusters as was previously observed in streptococcal (25) and lactococcal phages. In fact, in phage A2 most of this region could be deleted without any phenotypic effects, at least under laboratory conditions (49).

Pac-site *Siphoviridae*: ϕg1e, LL-H

The *Lactobacillus* phages ϕg1e and LL-H rely on *pac* sites for headful DNA packaging that results in variable genome segments of somewhat more than unit length. Their head gene cluster differs from that of the *cos*-site phages by encoding an additional head protein and a separate scaffold protein.

The overall genomic organization of the temperate *L. plantarum* phage ϕg1e (46) (figure 41-2B) was identical to that of *pac*-site *Streptococcus thermophilus* phage O1205, *L. lactis* phage TP901-1, *B. subtilis* phage SPP1 (9), *Listeria monocytogenes* phage A118 (51) and *Staphylococcus aureus* prophages ETA, all members of the proposed genus of ϕSfi11-like *pac*-site *Siphoviridae*. Frequently, even the gene order within an individual module was remarkably well

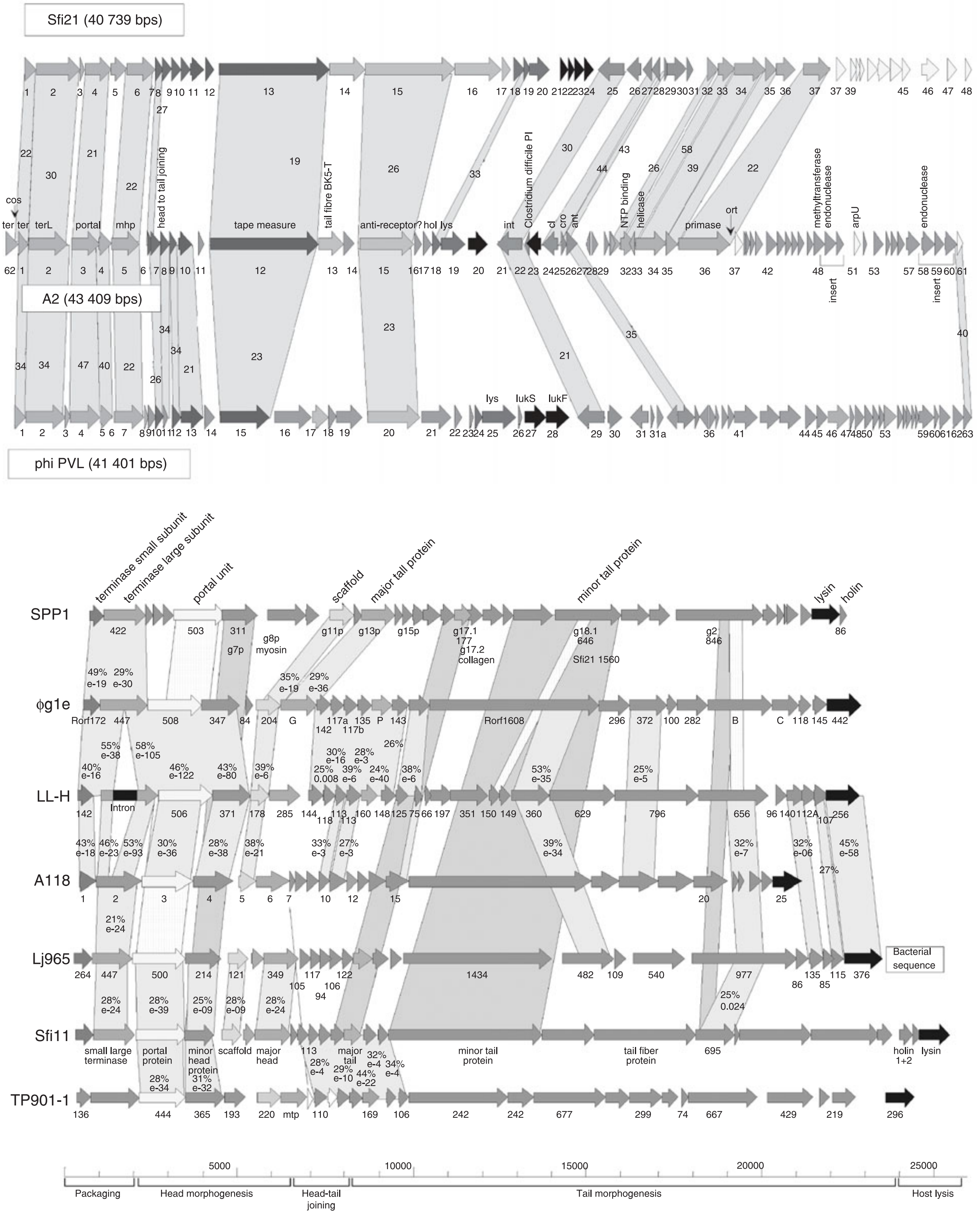

Sfi21 (40 739 bps)
cos
ter ter terL portal mhp head to tail joining tape measure tail fibre BK5-T anti-receptor? hol lys int Clostridium difficile PI cI cro ant NTP binding helicase primase ort methyltransferase endonuclease arpU endonuclease
insert
insert
A2 (43 409 bps)
lys lukS lukF
phi PVL (41 401 bps)
terminase small subunit
terminase large subunit
portal unit
scaffold
major tail protein
minor tail protein
lysin
holin
SPP1
g7p
g8p myosin
g11p g13p g15p
g17.1 177
g17.2 collagen
g18.1 646
Sfi21 1560
g2 846
ϕg1e
Rorf172
Rorf1608
LL-H
Intron
A118
Lj965
Bacterial sequence
Sfi11
small large terminase
portal protein
minor head protein
scaffold
major head
major tail
minor tail protein
tail fiber protein
holin 1+2
lysin
TP901-1
mtp
5000
10000
15000
20000
25000
Packaging
Head morphogenesis
Head-tail joining
Tail morphogenesis
Host lysis

Figure 41-2 Comparative genome maps of selected *Lactobacillus* phages. Top: Alignments of the genetic maps from the temperate *cos*-site *Streptococcus thermophilus* phage Sfi21, *L. casei* phage A2, and *Staphylococcus aureus* phage PVL. Genes that share significant sequence identity at the protein level are linked by gray shading; the bold figures indicate the degree of amino acid identity in % Bottom: Alignment of the late gene cluster from *pac*-site *Siphoviridae* from *Bacillus subtilis* phage SPP1, *L. plantarum* phage ϕg1e, *L. delbrueckii* phage LL-H, *Listeria monocytogenes* phage A118, *L. johnsonii* prophage Lj965, *Streptococcus thermophilus* phage Sfi11, and *Lactococcus lactis* phage TP901-1. Genes sharing significant sequence identity at the protein level are linked by gray (neighbors) or light- or dark-gray shading (for non-adjacent maps); the percentages correspond to amino acid identity and E-value. See thebacteriophages.org/frames_0410.htm for a color version of this figure.

conserved across species and genus barriers (lysogeny module, DNA replication module, structural genes). *L. delbrueckii* phage LL-H and *L. plantarum* phage ϕg1e shared related DNA packaging and head and tail morphogenesis modules. However, the sequence similarity was restricted to the protein level (suggesting that they are only distant relatives). Notably, the sequence similarity did not include the major head gene (39), demonstrating that individual genes and not entire modules are the units of genetic exchange between these *Lactobacillus* phages (22).

In contrast, high levels of DNA sequence identity were demonstrated between the virulent and the temperate *pac*-site phages LL-H and mv4 that infect the same host species *L. delbrueckii* (76). Notably, the virulent phage LL-H still contained remnants of a phage integrase and a phage attachment site, demonstrating that it was recently derived from a temperate phage (58). Close genetic relationships between temperate and virulent phages were also described in other dairy bacteria (56).

Analysis of *Lactobacillus* Phage Functions

Receptors

The pioneering work of Watanabe et al. (81–84) on the recognition and entry of phage PL-1 into its *L. casei* host indicated that the process can be divided into two steps: adsorption and injection. The first one is reversible, being the specific recognition between the phage and its host mediated by a cell wall polymer that contains rhamnose as a dominant residue. DNA injection is probably dependent on a membrane protein analogous to the phage infection protein (PIP) used by the *L. lactis* prolate-head phages of the c2 quasi-species (31). On the phage side, a protein, probably located at the tip of the LL-H virion fiber (gp71), has been found to recognize the receptors of its host, *L. delbrueckii* (68). A series of point mutations in the C-terminal part of gp71 widen the host range of the phage. Significantly, gp71 presented a significant degree of similarity to a protein from phage JCL1032, which infects the same host but is otherwise not related to LL-H.

Lysogeny Module

The determinants that encode the temperate phenotype, namely the genetic switch and the integration cassette, are clustered in all *Siphoviridae* able to lysogenize members of the low G-C branch of Gram-positive bacteria (55). The module is typically located between the lysis and replication cassettes and the genes are divergently transcribed. To one side of the genetic switch, the first gene encodes the lysogenic cycle repressor (*cI*, following the λ nomenclature), which is followed by two *orf*s of unknown function. The final gene to this side encodes the site-specific phage recombinase that mediates insertion of the phage DNA into the genome of its host. The recombinase is followed by the phage attachment (*attP*) region, where the recombination with the homologous *attB* site of the host chromosome takes place. In the vegetative phage map these genes oppose the convergently oriented lysis genes, being separated by a transcription termination loop.

Genetic Switch

To the other side of *cI* lies a *cro*-like repressor gene that may be followed by a putative antirepressor gene (reviewed for phage λ in chapter 8). Between *cI* and *cro* there is an intergenic region that harbors the two P_L and P_R promoters and the dyad-symmetry operator sequences for CI and Cro repressor binding (figure 41-3). The functionality of the CI-like repressors has been proven for three phages — ϕadh, ϕg1e and A2 — on the basis of their capacity to confer superinfection immunity to a previously susceptible host and by their ability to suppress transcription from the P_L and P_R promoters (24, 28, 46, 49). The CI repressors from phages ϕg1e and A2 have been purified and shown to bind in a cooperative manner their operator sequences. In both cases and as expected, the affinity of CI was higher for the operator sequences located in the vicinity of the P_R promoter. In addition, the CI repressor of phage A2 reallocated the RNA polymerase toward P_L, in spite of its much stronger affinity for P_R in the absence of any regulatory protein, thus promoting transcription of *cI* and the entry of the phage into the lysogenic cycle. For both phages, as the levels of CI were raised, all the operator sequences became occupied, resulting in autoregulation of CI biosynthesis (28, 39). Conversely, the purified Cro-like repressors from phages ϕg1e and A2 initially bound the operator sequences lying between the −35 and −10 hexamers of P_L and, as their concentrations increased, they also attached to the P_R overlapping operators (38, 40, 47). In the case of phage A2 this resulted in DNA-looping and displacement of the RNA polymerase from it (48).

Both CI and Cro repressors are dimers in solution and present DNA-binding helix–turn–helix motifs in their N-terminal moiety. Their C-terminal ends, by contrast, are responsible for dimerization. In the case of phage A2, the C-terminal end in addition enhances the affinity of Cro for the DNA (38, 47).

Regulation of Phage Life Cycles

The general features of the lysis/lysogeny decision of *Lactobacillus* phages resemble those of the λ genetic switch (35, 67), but their structural characteristics and regulation seem to be simpler. First, the synthesis of CI is constitutive and probably coexists with the expression of the lytic operon from the start of infection (40, 49). The situation may be facilitated (i) by the long intergenic region between

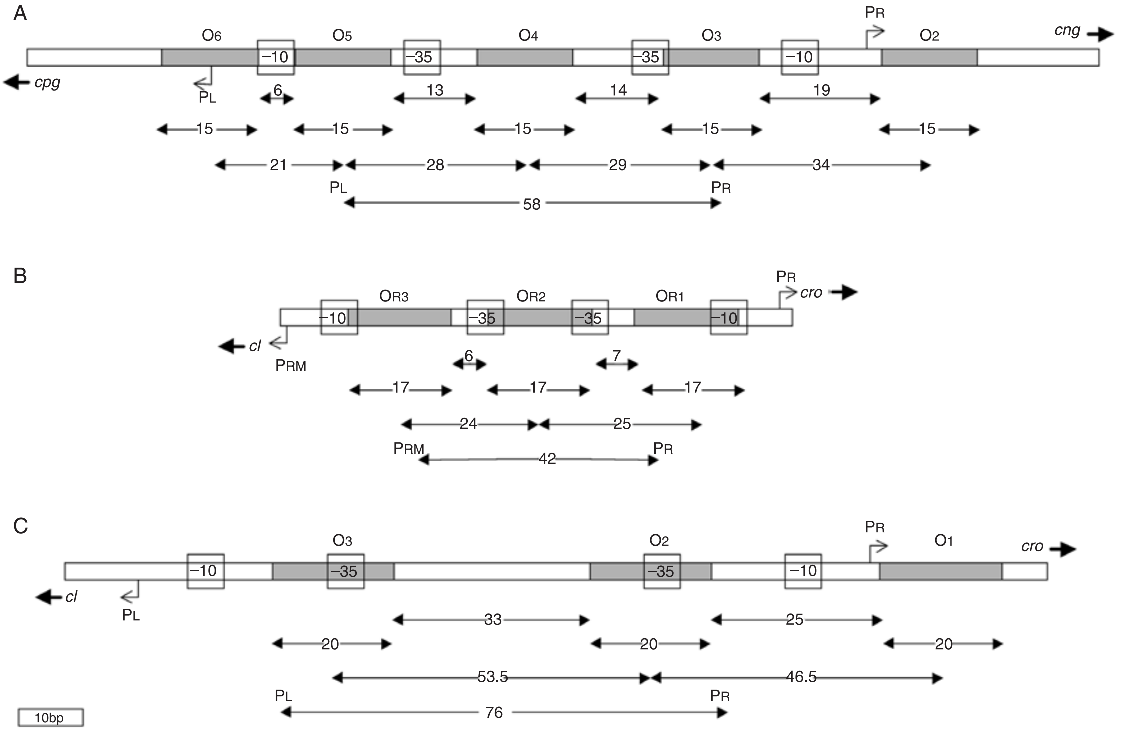


Figure 41-3 Comparison of the genetic switch regions. All are drawn to scale. A: Phages ϕg1e. B: Phage λ. C: Phage A2. The operator sequences (O_1 to O_6) are enclosed in shaded boxes. The −10 and −35 promoter consensus regions are indicated by open boxes. The transcription start sites of the divergently oriented promoters are signaled by bent arrows. The numbers within the double-headed arrows indicate the lengths (in base pairs) of the relevant features as follows (from top to bottom): distance between operator sequences, operators, segment from center to center of contiguous operators, stretch between the centers of the P_L and P_R promoters; note especially the difference in size of the segments located between the operator sites and the total length of the P_L–P_R intervening sequence. In the case of ϕg1e another operator sequence was detected 3′ of each *cpg* and *cng* (which are the functional equivalents to phage λ *cI* and *cro* respectively).

cI and *cro*, where two RNA polymerases may coexist (as has been shown in vitro for phage A2; 28) and (ii) by the lack of overlap between P_L and P_R (in the case of phage ϕadh, the situation may be different because the two promoters overlap; 24). In order to direct a significant proportion of the infectious events toward the lytic route, the constitutive synthesis of CI presumably has to be counterbalanced by an efficient production of Cro. This might explain why, in the case of phage A2, the transcripts arising from P_R are at least 10 times more abundant than those generated from P_L (28). In addition, the P_R transcript, which covers *cro*, elongates further into the early region to comprise the replication genes in the absence of any phage-encoded products (such as the λ N antitermination protein) (59 and unpublished data).

Integration

The integration systems of several *Lactobacillus* phages, including ϕFSW of *L. casei* (74, 75), ϕadh of *L. gasseri* (69), mv4 of *L. delbrueckii* (23), ϕg1e of *L. plantarum* (85) and A2 of *L. casei* (2), have been functionally characterized and various food-grade integration vectors were constructed (57, 73). The five recombinases belong to the tyrosine integrase family (4, 62), ranging in size from 385 residues (phages ϕadh and A2) to 427 residues (phage mv4). The prophages integrate into the 3′ end of tRNA genes, except for phage ϕFSW, in which integration occurs at the end of the glucose-6-phosphate isomerase gene. In all cases, their functions are conserved. Only Lj965, a prophage of *Lactobacillus johnsonii*, disrupts a $tRNA^{pro}$ gene (curiously this prophage carries four tRNA genes but none is specific for proline) (79).

The length of the common *attP*/*attB* core is between 16 bp (ϕadh) and 40 bp (ϕFSW). The recognition requirements of *attB* seem not to be very strict. Both the phage mv4 and phage A2 integrases promote integration of plasmids carrying their cognate *attP* DNA sequences into the genomes of a variety of Gram-positive bacteria of low G-C content and even into *E. coli* (2, 6, 7). Plasmid insertion takes place in positions homologous to their natural *attB* sequences in lactobacilli but it occurs in a variety of places, whether protein coding or intergenic, in other species. This variety of *attB*-like sites allowed the definition of minimum sequences for the integration to occur. In the case of phage mv4, this was a 16 bp DNA segment, where some nucleotides were critical for integration while others could be changed without loss of recombination efficiency (6). In phage A2 a conserved heptanucleotide, embedded in a 19 bp degenerated sequence, defined the *attB* site (2). All *attP* regions showed a high proportion of AT sequence and several direct and inverse repeats. These probably act as recognition sites for the integrases in order to generate the intasome complex that precedes synapsis formation with the *attB* site (71). The minimal size of *attP* has only been determined for phage mv4. It comprises a DNA segment slightly longer than 200 bp. The *attP*-core sequence is found at one of its ends (5).

Replication

Lactobacillus phage replication functions have not been investigated in great detail. No replication enzyme has been isolated and the only functional studies deal with the replication origins (*ori*) of phages ϕadh and A2. The *ori* region, acting in *cis*, promotes plasmid replication in their natural hosts and, in the case of phage ϕadh, also in *L. lactis* (1). In phage A2, *ori* may act as a decoy for phage-specific proteins because, when present in multicopy, it partially inhibited phage replication. This resulted in a significantly lower phage progeny count and thus in partial resistance against infection. Moreover, a DNA fragment containing *ori* was retarded in gel mobility-shift experiments when incubated with extracts of infected cells (59).

Morphogenesis

Information on the processes leading to virion formation in lactobacilli mainly refers to phages ϕg1e and A2. In phage ϕg1e a major structural protein was purified and specific antibodies were raised that allowed its ascription to the tail of the virion (41). In phage A2, the virion presents three major proteins, one of which was allocated to the tail while the other two were capsid components. These two shared their N-termini (29). Maldi-Tof mass spectrometry indicated that the smaller (gp5A) is the product of gene 5 translation. Its large counterpart (gp5B) is originated through a −1 ribosomal frameshift at the penultimate codon of *orf*5 mRNA, resulting in a product that is 85 amino acids longer than gp5A (30). This is similar to the situation described for several *E. coli* phages, exemplified by T3 and T7, where gene *10* yields two major coat proteins (17, 18) (see chapter 20 for a review of phage T7, and T3, biology). Both gp5A and gp5B appear to be essential for phage viability because lysogens harboring prophages that produce only one or the other protein become lysed upon induction with mitomycin C but no viable phage progeny are observed (30).

Additional information on the morphogenesis of *Lactobacillus* phages comes from phage A2, whose terminase small subunit has been characterized. The protein specifically recognizes a short DNA segment that includes the *cos* region of the phage. In an ATP-dependent reaction, the terminase small subunit induces bending in the *cos* region prior to DNA staggered cleavage at *cosN* (27).

Lysis

The lysis modules of six different phages (mv1, ϕadh, ϕg1e, PL-1, LL-H, and SC921) had been investigated (11, 34, 42, 43, 77, 87) due to their interest for accelerated food ripening (19, 72). However, the first cell wall lytic activity was purified from PL-1-infected cultures of *L. casei*, with the aim of using it in protoplast generation (as a prerequisite for DNA transformation). The enzyme turned out to be an endo-*N*-acetylmuramidase of 37 kDa, mostly active on the cell wall of the phage's host (80). However, Kashige et al., (43) reported the sequence of a PL-1-encoded acetylmuramoyl-L-alanine amidase, casting doubt on the origin of the previously isolated lysozyme-like enzyme. The PL-1 amidase gene is included in a bicistronic operon and is preceded by a putative holin gene, the typical gene arrangement of the phage lysis cassettes in lactic acid bacteria. Curiously, this is the only amidase identified so far from phages of lactobacilli; all other investigated lysins are muramidases.

Structurally these muramidases are rather diverse, their sizes ranging from 195 amino acids (mv1) (11) up to 442 residues (ϕg1e) (34, 42, 64, 77). Usually they are most active on the walls of their host bacteria. For example, LysA from phage mv1 and Mur from phage LL-H efficiently degrade the peptidoglycan of thermophilic lactobacilli (their host, *L. delbrueckii*, belongs to this group) but are inactive on walls of *L. casei* or *L. lactis* (11, 77). In the case of the phage ϕg1e lysin, several acidic residues and a serine, all located in its N-terminal moiety, were essential for the lytic activity while deletion of the 45 residues nearest the C-terminal did not abrogate lysis induction.

A striking feature of the ϕg1e lysin is the presence of a signal peptide of 26 amino acids cleaved from the mature enzyme. This suggests that secretion of the lysin through the cytoplasmic membrane occurs via a Sec-A-dependent mechanism (42). However, the ϕg1e lysis cassette also harbors the gene for a functional holin, as determined by the formation of cell ghosts upon expression of this protein in *E. coli* (63). This poses the question on whether both lysin export systems might be functional in the *L. plantarum* host cells. The holin of phage ϕg1e, is a basic polypeptide of 142 residues that contains three hydrophobic domains in its N-terminal moiety which promote insertion of the protein in the cell membrane as a prerequisite for pore formation, which is also dependent on the C-terminal part of the protein (63). See chapter 10 for a review of holin-mediated phage-induced bacterial lysis.

Lactobacillus Prophages

Prophages represent a sizable part of the strain-specific DNA in many bacteria (for recent reviews, see 15, 16a). Since phages are prominent mobile DNA elements, this observation was not unexpected. Two not necessarily exclusive hypotheses were proposed with respect to the evolutionary role of prophage DNA in bacterial genomes. The present prophage content could represent snapshots in the arms race between bacteria and phages where selfish DNA elements litter the bacterial genomes. In this model, prophages are constantly acquired by lysogeny and, since they represent a threat to the survival of the cell (prophage induction) and a metabolic burden (extra DNA for replication), they are constantly removed by selection using a nonspecific deletion process (50). In an alternative model, prophages play an active and sometimes mutualistic role for short-term bacterial evolution. In fact, prophages contribute a constant influx of extra genes (lysogenic conversion genes) into a given bacterial species and thus allow the selection of strains that are best adapted to a given environment (13). Prophage genomics in some pathogenic bacteria such as *Streptococcus pyogenes* supports the second model. They carry genes, expressed during lysogeny, that encode superantigens/toxins, mitogenic molecules/Dnases, and other virulence factors involved in disease (8, 10, 36).

Do prophages also play a role for the ecological adaptation of nonpathogenic microorganisms? To address this question, the genomes from two recently sequenced *Lactobacillus* commensals were investigated for their prophage content. The gut commensal *L. johnsonii* strain NCC533 contained two apparently complete, but noninducible prophages, Lj928 and Lj965, and a prophage remnant in addition to numerous isolated phage-like integrase genes (79). The two complete prophages were classified as ϕSfi11-like *pac*-site phages (12). However, sequence similarity between them was restricted to a few genes. The similarity included, surprisingly, the phage integrase that apparently obliged Lj965 to use a secondary attachment site leading to the inactivation of a tRNA gene (79). Both prophages contained extra genes of unknown function within the lysogeny module. Further extra genes was inserted upstream of the structural gene module and encoded several tRNAs. Most of the extra genes were constitutively transcribed in the lysogen while the rest of the prophage genome was transcriptionally almost silent. When the DNA from eight *L. johnsonii* strains showing distinct pulsed-field restriction patterns was hybridized to a microarray of the sequenced strain NCC533, it was observed that the prophage DNA contributed about half of the strain NCC533-specific DNA (79).

The oral *L. plantarum* isolate WCFS1 (45) contained two complete prophages, Lp1 and Lp2, also members of the ϕSfi11-like group of *pac*-site *Siphoviridae* (78). They shared long stretches of DNA sequence identity, including one of about 10 kb over their structural gene modules.

Phages Lp1 and Lp2 lacked sequence relatedness with the *L. plantarum* phage ϕg1e. This observation suggests the existence of two lineages of ϕSfi11-like *pac*-site *Siphoviridae* in *L. plantarum*. Genes without obvious phage links were located by comparative genomics analysis in the lysogeny module and between the lysin gene and the right attachment site. Notably, two of these genes shared sequence similarity with likely lysogenic conversion genes from *Streptococcus pyogenes* prophages (including mitogenic factor-like genes). This extra DNA, which again included tRNA genes, belonged to the few transcribed segments of the prophage genome. WCFS1 also contained two prophage remnants. Remnant R-Lp3 consisted of a few genes from the lysogeny, DNA replication, head and head-to-tail joining genes of a typical Sfi21-like *cos*-site phage, but it lacked sequence similarities with phages ϕadh and A2, suggesting the existence of a second lineage of *cos*-site *Siphoviridae* in *Lactobacillus*. R-Lp3 suffered extensive gene losses, fusion of gene fragments, and DNA rearrangements. R-Lp3 abutted Lp2 and showed no transcriptional activity. Another short prophage remnant, R-Lp4, showed incomplete lysogeny and DNA replication modules and a gene with a database link to an anonymous gene from an *Enterococcus faecalis* pathogenicity island.

The observations with *Lactobacillus* prophages concur with recent predictions on the role of prophages for the genome evolution of the bacterial hosts. Elements of an arms race between phages and their hosts are apparent: the bacterial counterattack is, for example, suggested by the apparent inactivation of prophages, and the presence of prophage remnants and many isolated phage-like integrase genes. At the same time, prophage analysis suggests signs of genetic cooperation. For example, there is a constitutive transcription of candidate lysogenic conversion genes (14).

Ecology

Lactobacillus bacteriophages are of substantial research interest since their host bacteria are involved in numerous industrial food fermentation processes. In the dairy industry, *L. delbrueckii* is used in a coculture with *Streptococcus thermophilus* for the production of yogurt. While phage attack is a serious problem for *S. thermophilus*, *L. delbrueckii* is a relatively rare target of phage attack in industrial fermentations, suggesting that some kind of defense against phage infection might be operative in these strains.

Sauerkraut fermentation is spontaneous and relies on bacterial epiphytes present on cabbage. An ecological survey demonstrated a succession of two phage populations corresponding to the replacement of *Leuconostoc* spp. by *Lactobacillus* spp. within the fermenting vats (53, 86).

Lactobacillus phages have been isolated from numerous other spontaneous fermentation processes involving coffee, pickled cucumbers, cereals, wine, and meat (52, 61). However, phage infections during salami production had no industrial impact. After an initial rise, the phage titers dropped and the initial phage-sensitive starter population was replaced by a phage-insensitive mutant derivative strain (60).

In addition, lactobacilli are normal inhabitants of the alimentary and urogenital tract of man and many animals. Lactobacilli were therefore proposed as probiotic, i.e., health-promoting bacteria (70). The production of probiotic strains can be delicate since they represent monocultures that may be especially susceptible to phage attack. In practice, some were produced for long periods in the absence of phage problems. A possible reason for this phage resistance might be their polylysogenic nature.

However, phage attack on *Lactobacillus* commensals was suggested to have health consequences (44). Lactobacilli constitute the dominant bacterial microbiota in the vagina and are beneficial to women's health since they inhibit the growth of harmful microorganisms. About 50% of women with bacterial vaginosis carry lysogenic lactobacilli in their vagina. The mutagen benzopyrene created by tobacco smoking induced phages from lysogenic lactobacilli at concentrations that were found in vaginal secretions of women who smoked (65). This led to the intriguing scenario that smoking might reduce vaginal lactobacilli by promoting phage induction, which would lead to an overgrowth by anaerobic bacteria and thus to bacterial vaginosis.

For the genomics researcher, *Lactobacillus* phages are an interesting group for two reasons. First, sequencing data are available for phages from five distinct species of lactobacilli (*L. delbrueckii, gasseri, plantarum, casei,* and *johnsonii*). Second, in contrast to most phages that show narrow host ranges (however, see chapter 46), several *Lactobacillus* phages can infect two or more members of the genus. One sauerkraut phage isolate infected ecologically related *L. brevis* and *L. plantarum* (53) strains and inducible prophages from vagina showed a wide host range including *L. crispatus, jensenii, gasseri, fermentum,* and *vaginalis* (44). This property offers substantial opportunities for phage-mediated lateral gene transfer between different lactobacilli.

Outlook

Research in the field of *Lactobacillus* phages has been driven by very different motives leading to a situation where some data are available for many phages but no comprehensive dataset exists for a type phage. This is a clear disadvantage for classical phage biologists reared on the diet of reductionist thinking. However, phage biology is currently at a turning point from serving as simple model systems in molecular biology to providing handy tools for the exploration of complex ecological relationships. In this respect, *Lactobacillus* phages have something to offer: they comprise many different phage types and infect hosts that are found in many and diverse ecological settings (several of them with clear economic and medical interest). Two assets can still be quoted from this chapter's introduction. First, in contrast to lactococci and lactic streptococci, there exists an extensive literature on the roles played by lactobacilli in their different habitats. Second, the genome sequences from more than 10 different *Lactobacillus* species will soon be available. What we need now is a comparable sequencing effort for *Lactobacillus* phages and the stage will be set for a genomics-oriented molecular ecology of phage–host interactions on the mucosal surfaces of humans (alimentary and genital tract) or on plants. *Lactobacillus* phage research could thus become a meeting point for scientists currently working in separate fields (dairy science, food microbiology, medical microbiology, ecology). As an example, dairy microbiologists have extensively explored the possibilities for protecting starter bacteria against phage attack, while medical microbiologists have started to seriously investigate the killing potential of lytic phages on bacterial pathogens. What about cross-fertilizing the activities between dairy and medical microbiologists by developing phage-resistant probiotic lactobacilli?

Acknowledgments

The research in H.B.'s laboratory was supported by two successive grants from the Swiss National Science Foundation. The work in the J. E. S.'s laboratory is financed by grants SAF2004-0033 and BMC2002-0638 from CICYT (cosponsored by the Plan FEDER from the European Union). We also thank J. L. Caso and M.-L. Dillmann for the A2 and the ΦLp65 micrographs, respectively.

References

1. Altermann, E., J. R. Klein, and B. Henrich. 1999. Primary structure and features of the genome of the *Lactobacillus gasseri* temperate bacteriophage ϕadh. Gene 236:333–346.
2. Alvarez, M. A., M. Herrero, and J. E. Suarez. 1998. The site-specific recombination system of the *Lactobacillus* species bacteriophage A2 integrates in gram-positive and gram-negative bacteria. Virology 250:185–193.
3. Alvarez, M. A., A. Rodriguez, and J. E. Suarez. 1999. Stable expression of the *Lactobacillus casei* bacteriophage A2 repressor blocks phage propagation during milk fermentation. J. Appl. Microbiol. 86:812–816.
4. Argos, P., A. Landy, K. Abremski, J. B. Egan, E. Haggard-Ljungquist, R. H. Hoess, M. L. Kahn, B. Kalionis,

S. V. Narayana, and L. S. Pierson, III. 1986. The integrase family of site-specific recombinases: regional similarities and global diversity. EMBO J. 5:433–440.

5. Auvray, F., M. Coddeville, G. Espagno, and P. Ritzenthaler. 1999. Integrative recombination of *Lactobacillus delbrueckii* bacteriophage mv4: functional analysis of the reaction and structure of the *attP* site. Mol. Gen. Genet. 262:355–366.

6. Auvray, F., M. Coddeville, R. C. Ordonez, and P. Ritzenthaler. 1999. Unusual structure of the *attB* site of the site-specific recombination system of *Lactobacillus delbrueckii* bacteriophage mv4. J. Bacteriol. 181:7385–7389.

7. Auvray, F., M. Coddeville, P. Ritzenthaler, and L. Dupont. 1997. Plasmid integration in a wide range of bacteria mediated by the integrase of *Lactobacillus delbrueckii* bacteriophage mv4. J. Bacteriol. 179:1837–1845.

8. Banks, D. J., S. B. Beres, and J. M. Musser. 2002. The fundamental contribution of phages to GAS evolution, genome diversification and strain emergence. Trends Microbiol. 10:515–521.

9. Becker, B., F. N. de la Fuente, M. Gassel, D. Gunther, P. Tavares, R. Lurz, T. A. Trautner, and J. C. Alonso. 1997. Head morphogenesis genes of the *Bacillus subtilis* bacteriophage SPP1. J. Mol. Biol. 268:822–839.

10. Beres, S. B., G. L. Sylva, K. D. Barbian, B. Lei, J. S. Hoff, N. D. Mammarella, M. Y. Liu, J. C. Smoot, S. F. Porcella, L. D. Parkins, D. S. Campbell, T. M. Smith, J. K. McCormick, D. Y. Leung, P. M. Schlievert, and J. M. Musser. 2002. Genome sequence of a serotype M3 strain of group A *Streptococcus*: phage-encoded toxins, the high-virulence phenotype, and clone emergence. Proc. Natl. Acad. Sci. USA 99:10078–10083.

11. Boizet, B., Y. Lahbib-Mansais, L. Dupont, P. Ritzenthaler, and M. Mata. 1990. Cloning, expression and sequence analysis of an endolysin-encoding gene of *Lactobacillus bulgaricus* bacteriophage mv1. Gene 94:61–67.

12. Brüssow, H., and F. Desiere. 2001. Comparative phage genomics and the evolution of *Siphoviridae*: insights from dairy phages. Mol. Microbiol. 39:213–222.

13. Canchaya, C., F. Desiere, W. M. McShan, J. J. Ferretti, J. Parkhill, and H. Brüssow. 2002. Genome analysis of an inducible prophage and prophage remnants integrated in the *Streptococcus pyogenes* strain SF370. Virology 302:245–258.

14. Canchaya, C., C. Proux, G. Fournous, A. Bruttin, and H. Brüssow. 2003. Prophage genomics. Microbiol. Mol. Biol. Rev. 7:238–276, table.

15. Canchaya, C., C. Proux, G. Fournous, A. Bruttin, and H. Brüssow. 2003. Prophage genomics. Microbiol. Mol. Biol. Rev. 67:238–276 + erratumn on page 473.

16a. Casjens, S. 2003. Prophages and bacterial genomics: what have we learned so far? Mol. Microbiol. 49:277–300.

16b. Chibani-Chennoufi S., M. L. Dillmann, L. Marvin-Guy, S. Rami-Shojaei, and H. Brüssow. 2004. *Lactobacillus plantarum* bacteriophage LP65: a new member of the SP01-like genus of the family Myoviridae. J. Bacteriol. 186:7069–7083.

17. Condreay, J. P., S. E. Wright, and I. J. Molineux. 1989. Nucleotide sequence and complementation studies of the gene 10 region of bacteriophage T3. J. Mol. Biol. 207:555–561.

18. Condron, B. G., J. F. Atkins, and R. F. Gesteland. 1991. Frameshifting in gene 10 of bacteriophage T7. J. Bacteriol. 173:6998–7003.

19. de Ruyter, P. G., O. P. Kuipers, W. C. Meijer, and W. M. de Vos. 1997. Food-grade controlled lysis of *Lactococcus lactis* for accelerated cheese ripening. Nat. Biotechnol. 15:976–979.

20. Desiere, F., S. Lucchini, and H. Brüssow. 1999. Comparative sequence analysis of the DNA packaging, head, and tail morphogenesis modules in the temperate *cos*-site *Streptococcus thermophilus* bacteriophage Sfi21. Virology 260:244–253.

21. Desiere, F., C. Mahanivong, A. J. Hillier, P. S. Chandry, B. E. Davidson, and H. Brüssow. 2001. Comparative genomics of lactococcal phages: insight from the complete genome sequence of *Lactococcus lactis* phage BK5-T. Virology 283:240–252.

22. Desiere, F., R. D. Pridmore, and H. Brüssow. 2000. Comparative genomics of the late gene cluster from *Lactobacillus* phages. Virology 275:294–305.

23. Dupont, L., B. Boizet-Bonhoure, M. Coddeville, F. Auvray, and P. Ritzenthaler. 1995. Characterization of genetic elements required for site-specific integration of *Lactobacillus delbrueckii* subsp. *bulgaricus* bacteriophage mv4 and construction of an integration-proficient vector for *Lactobacillus plantarum*. J. Bacteriol. 177: 586–595.

24. Engel, G., E. Altermann, J. R. Klein, and B. Henrich. 1998. Structure of a genome region of the *Lactobacillus gasseri* temperate phage ϕadh covering a repressor gene and cognate promoters. Gene 210:61–70.

25. Foley, S., A. Bruttin, and H. Brüssow. 2000. Widespread distribution of a group I intron and its three deletion derivatives in the lysin gene of *Streptococcus thermophilus* bacteriophages. J. Virol. 74:611–618.

26. Forsman, P. 1993. Characterization of a prolate-headed bacteriophage of *Lactobacillus delbrueckii* subsp. *lactis*, and its DNA homology with isometric-headed phages. Arch. Virol. 132:321–330.

27. Garcia, P., J. C. Alonso, and J. E. Suarez. 1997. Molecular analysis of the cos region of the *Lactobacillus casei* bacteriophage A2. Gene product 3, gp3, specifically binds to its downstream *cos* region. Mol. Microbiol. 23:505–514.

28. Garcia, P., V. Ladero, J. C. Alonso, and J. E. Suarez. 1999. Cooperative interaction of CI protein regulates lysogeny of *Lactobacillus casei* by bacteriophage A2. J. Virol. 73:3920–3929.

29. Garcia, P., V. Ladero, and J. E. Suarez. 2003. Analysis of the morphogenetic cluster and genome of the temperate *Lactobacillus casei* bacteriophage A2. Arch. Virol. 148:1051–1070.

30. Garcia, P., I. Rodriguez, and J. E. Suarez. 2004. A −1 ribosomal frameshift in the transcript that encodes the major head protein of bacteriophage A2 mediates biosynthesis of a second essential component of the capsid. J. Bacteriol. 186:1714–1719.

31. Geller, B. L., R. G. Ivey, J. E. Trempy, and B. Hettinger-Smith. 1993. Cloning of a chromosomal gene required for phage infection of *Lactococcus lactis* subsp. *lactis* C2. J. Bacteriol. 175:5510–5519.

32. Gilakjan, Z. A., and A. M. Kropinski. 1999. Cloning and analysis of the capsid morphogenesis genes of *Pseudomonas aeruginosa* bacteriophage D3: another example of protein chain mail? J. Bacteriol. 181:7221–7227.
33. Hendrix, R. W., and R. L. Duda. 1998. Bacteriophage HK97 head assembly: a protein ballet. Adv. Virus Res. 50:235–288.
34. Henrich, B., B. Binishofer, and U. Bläsi. 1995. Primary structure and functional analysis of the lysis genes of *Lactobacillus gasseri* bacteriophage ϕadh. J. Bacteriol. 177:723–732.
35. Hochschild, A., J. Douhan, III, and M. Ptashne. 1986. How lambda repressor and lambda Cro distinguish between OR1 and OR3. Cell 47:807–816.
36. Ikebe, T., A. Wada, Y. Inagaki, K. Sugama, R. Suzuki, D. Tanaka, A. Tamaru, Y. Fujinaga, Y. Abe, Y. Shimizu, and H. Watanabe. 2002. Dissemination of the phage-associated novel superantigen gene *speL* in recent invasive and noninvasive *Streptococcus pyogenes* M3/T3 isolates in Japan. Infect. Immun. 70:3227–3233.
37. Jarvis, A. W., L. J. Collins, and H. W. Ackermann. 1993. A study of five bacteriophages of the *Myoviridae* family which replicate on different gram-positive bacteria. Arch. Virol. 133:75–84.
38. Kakikawa, M., S. Ohkubo, T. Sakate, M. Sayama, A. Taketo, and K. Kodaira. 2000. Purification and DNA-binding properties of the cro-type regulatory repressor protein cng encoded by the *Lactobacillus plantarum* phage ϕg1e. Gene 249:161–169.
39. Kakikawa, M., S. Ohkubo, M. Syama, A. Taketo, and K. I. Kodaira. 2000. The genetic switch for the regulatory pathway of *Lactobacillus plantarum* phage ϕg1e: characterization of the promoter *P(L)*, the repressor gene *cpg*, and the *cpg*-encoded protein Cpg in *Escherichia coli*. Gene 242:155–166.
40. Kakikawa, M., N. Watanabe, T. Funawatashi, M. Oki, H. Yasukawa, A. Taketo, and K. I. Kodaira. 1998. Promoter/repressor system of *Lactobacillus plantarum* phage ϕg1e: characterization of the promoters *pR49–pR–pL* and overproduction of the cro-like protein cng in *Escherichia coli*. Gene 215:371–379.
41. Kakikawa, M., A. Yamakawa, K. J. Yokoi, S. Nakamura, A. Taketo, and K. Kodaira. 2002. Characterization of the major tail protein gpP encoded by *Lactobacillus plantarum* phage ϕg1e. J. Biochem. Mol. Biol. Biophys. 6:185–191.
42. Kakikawa, M., K. J. Yokoi, H. Kimoto, M. Nakano, K. Kawasaki, A. Taketo, and K. Kodaira. 2002. Molecular analysis of the lysis protein Lys encoded by *Lactobacillus plantarum* phage ϕg1e. Gene 299:227–234.
43. Kashige, N., Y. Nakashima, F. Miake, and K. Watanabe. 2000. Cloning, sequence analysis, and expression of *Lactobacillus casei* phage PL-1 lysis genes. Arch. Virol. 145:1521–1534.
44. Kilic, A. O., S. I. Pavlova, S. Alpay, S. S. Kilic, and L. Tao. 2001. Comparative study of vaginal *Lactobacillus* phages isolated from women in the United States and Turkey: prevalence, morphology, host range, and DNA homology. Clin. Diagn. Lab. Immunol. 8:31–39.
45. Kleerebezem, M., J. Boekhorst, R. van Kranenburg, D. Molenaar, O. P. Kuipers, R. Leer, R. Tarchini, S. A. Peters, H. M. Sandbrink, M. W. Fiers, W. Stiekema, R. Lankhorst, P. A. Bron, S. M. Hoffer, M. N. Groot, R. Kerkhoven, M. de Vries, B. Ursing, W. M. de Vos, and R. J. Siezen. 2003. Complete genome sequence of *Lactobacillus plantarum* WCFS1. Proc. Natl. Acad. Sci. USA 100: 1990–1995.
46. Kodaira, K. I., M. Oki, M. Kakikawa, N. Watanabe, M. Hirakawa, K. Yamada, and A. Taketo. 1997. Genome structure of the *Lactobacillus* temperate phage ϕg1e: the whole genome sequence and the putative promoter/repressor system. Gene 187:45–53.
47. Ladero, V., P. Garcia, J. C. Alonso, and J. E. Suarez. 1999. A2 Cro, the lysogenic cycle repressor, specifically binds to the genetic switch region of *Lactobacillus casei* bacteriophage A2. Virology 262:220–229.
48. Ladero, V., P. Garcia, J. C. Alonso, and J. E. Suarez. 2002. Interaction of the Cro repressor with the lysis/lysogeny switch of the *Lactobacillus casei* temperate bacteriophage A2. J. Gen. Virol. 83:2891–2895.
49. Ladero, V., P. Garcia, V. Bascaran, M. Herrero, M. A. Alvarez, and J. E. Suarez. 1998. Identification of the repressor-encoding gene of the *Lactobacillus* bacteriophage A2. J. Bacteriol. 180:3474–3476.
50. Lawrence, J. G., R. W. Hendrix, and S. Casjens. 2001. Where are the pseudogenes in bacterial genomes? Trends Microbiol. 9:535–540.
51. Loessner, M. J., R. B. Inman, P. Lauer, and R. Calendar. 2000. Complete nucleotide sequence, molecular analysis and genome structure of bacteriophage A118 of *Listeria monocytogenes*: implications for phage evolution. Mol. Microbiol. 35:324–340.
52. Lu, Z., F. Breidt, Jr., H. P. Fleming, E. Altermann, and T. R. Klaenhammer. 2003. Isolation and characterization of a *Lactobacillus plantarum* bacteriophage, ϕJL-1, from a cucumber fermentation. Int. J. Food Microbiol. 84:225–235.
53. Lu, Z., F. Breidt, V. Plengvidhya, and H. P. Fleming. 2003. Bacteriophage ecology in commercial sauerkraut fermentations. Appl. Environ. Microbiol. 69:3192–3202.
54. Lucchini, S., F. Desiere, and H. Brüssow. 1998. The structural gene module in *Streptococcus thermophilus* bacteriophage ϕSfi11 shows a hierarchy of relatedness to *Siphoviridae* from a wide range of bacterial hosts. Virology 246:63–73.
55. Lucchini, S., F. Desiere, and H. Brüssow. 1999. Similarly organized lysogeny modules in temperate *Siphoviridae* from low GC content gram-positive bacteria. Virology 263:427–435.
56. Lucchini, S., F. Desiere, and H. Brüssow. 1999. The genetic relationship between virulent and temperate *Streptococcus thermophilus* bacteriophages: whole genome comparison of *cos*-site phages Sfi19 and Sfi21. Virology 260:232–243.
57. Martin, M. C., J. C. Alonso, J. E. Suarez, and M. A. Alvarez. 2000. Generation of food-grade recombinant lactic acid bacterium strains by site-specific recombination. Appl. Environ. Microbiol. 66:2599–2604.
58. Mikkonen, M., L. Raisanen, and T. Alatossava. 1996. The early gene region completes the nucleotide sequence of *Lactobacillus delbrueckii* subsp. *lactis* phage LL-H. Gene 175:49–57.

59. Moscoso, M., and J. E. Suarez. 2000. Characterization of the DNA replication module of bacteriophage A2 and use of its origin of replication as a defense against infection during milk fermentation by *Lactobacillus casei*. Virology 273:101–111.
60. Nes I. F., J. Brendehaug, and K. O. von Husby. 1988. Characterization of the bacteriophage B2 of *Lactobacillus plantarum* ATCC8019. Biochemie 70:423–427.
61. Nes I .F., and O. Sorheim. 1984. Effect of infection of a bacteriophage in a starter culture during the production of salami dry sausage: a model study. J. Food Sci. 49:337–340.
62. Nunes-Duby, S. E., H. J. Kwon, R. S. Tirumalai, T. Ellenberger, and A. Landy. 1998. Similarities and differences among 105 members of the Int family of site-specific recombinases. Nucleic Acids Res. 26:391–406.
63. Oki, M., M. Kakikawa, S. Nakamura, E. T. Yamamura, K. Watanabe, M. Sasamoto, A. Taketo, and K. Kodaira. 1997. Functional and structural features of the holin HOL protein of the *Lactobacillus plantarum* phage ϕg1e: analysis in *Escherichia coli* system. Gene 197:137–145.
64. Oki, M., M. Kakikawa, K. Yamada, A. Taketo, and K. I. Kodaira. 1996. Cloning, sequence analysis, and expression of the genes encoding lytic functions of bacteriophage ϕg1e. Gene 176:215–223.
65. Pavlova, S. I., and L. Tao. 2000. Induction of vaginal *Lactobacillus* phages by the cigarette smoke chemical benzo[a]-pyrene diol epoxide. Mutat. Res. 466:57–62.
66. Proux, C., D. van Sinderen, J. Suarez, P. Garcia, V. Ladero, G. F. Fitzgerald, F. Desiere, and H. Brüssow. 2002. The dilemma of phage taxonomy illustrated by comparative genomics of sfi21-like *Siphoviridae* in lactic acid bacteria. J. Bacteriol. 184:6026–6036.
67. Ptashne, M. 2003. A Genetic Switch. Cell Press and Blackwell Scientific, Oxford.
68. Ravin, V., L. Raisanen, and T. Alatossava. 2002. A conserved C-terminal region in Gp71 of the small isometric-head phage LL-H and ORF474 of the prolate-head phage JCL1032 is implicated in specificity of adsorption of phage to its host, *Lactobacillus delbrueckii*. J. Bacteriol. 184:2455–2459.
69. Raya, R. R., C. Fremaux, G. L. De Antoni, and T. R. Klaenhammer. 1992. Site-specific integration of the temperate bacteriophage ϕadh into the *Lactobacillus gasseri* chromosome and molecular characterization of the phage (*attP*) and bacterial (*attB*) attachment sites. J. Bacteriol. 174:5584–5592.
70. Redondo-Lopez, V., R. L. Cook, and J. D. Sobel. 1990. Emerging role of lactobacilli in the control and maintenance of the vaginal bacterial microflora. Rev. Infect. Dis. 12:856–872.
71. Richet, E., P. Abcarian, and H. A. Nash. 1988. Synapsis of attachment sites during lambda integrative recombination involves capture of a naked DNA by a protein-DNA complex. Cell 52:9–17.
72. Shearman, C. A. J. K. a. M. G. 2003. Autolytic *Lactococcus lactis* expressing a lactococcal bacteriophage lysin gene. Biotechnology 196–199.
73. Shimizu-Kadota, M. 2001. A method to maintain introduced DNA sequences stably and safely on the bacterial chromosome: application of prophage integration and subsequent designed excision. J. Biotechnol. 89:73–79.
74. Shimizu-Kadota, M., M. Kiwaki, S. Sawaki, Y. Shirasawa, H. Shibahara-Sone, and T. Sako. 2000. Insertion of bacteriophage ϕFSW into the chromosome of *Lactobacillus casei* strain Shirota (S-1): characterization of the attachment sites and the integrase gene. Gene 249:127–134.
75. Shimizu-Kadota, M. and N. Tsuchida. 1984. Physical mapping of the virion and the prophage DNAs of a temperate *Lactobacillus* phage ϕFSW. J. Gen. Microbiol. 130 (Pt 2): 423–430.
76. Vasala, A., L. Dupont, M. Baumann, P. Ritzenthaler, and T. Alatossava. 1993. Molecular comparison of the structural proteins encoding gene clusters of two related *Lactobacillus delbrueckii* bacteriophages. J. Virol. 67:3061–3068.
77. Vasala, A., M. Valkkila, J. Caldentey, and T. Alatossava. 1995. Genetic and biochemical characterization of the *Lactobacillus delbrueckii* subsp. *lactis* bacteriophage LL-H lysin. Appl. Environ. Microbiol. 61:4004–4011.
78. Ventura, M., C. Canchaya, M. Kleerebezem, W. M. de Vos, R. J. Siezen, and H. Brüssow. The prophages sequences of *Lactobacillus plantarum* strain WCFS1. Virology 316: 245–255.
79. Ventura, M., C. Canchaya, D. Pridmore, B. Berger, and H. Brüssow. 2003. Integration and distribution of *Lactobacillus johnsonii* prophages. J. Bacteriol. 185:4603–4608.
80. Watanabe, K., M. Hayashida, K. Ishibashi, and Y. Nakashima. 1984. An N–acetylmuramidase induced by PL-1 phage infection of *Lactobacillus casei*. J. Gen. Microbiol. 130:275–277.
81. Watanabe, K., K. Ishibashi, Y. Nakashima, and T. Sakurai. 1984. A phage-resistant mutant of *Lactobacillus casei* which permits phage adsorption but not genome injection. J. Gen. Virol. 65:981–986.
82. Watanabe, K., Y. Nakashima, and S. Kamiya. 1992. Effects of some L-rhamnosyl derivatives on the adsorption of phage PL–1 to the host *Lactobacillus casei*. Biosci. Biotechnol. Biochem. 56:346.
83. Watanabe, K., M. Shirabe, Y. Nakashima, and Y. Kakita. 1991. The possible involvement of protein synthesis in the injection of PL-1 phage genome into its host, *Lactobacillus casei*. J. Gen. Microbiol. 137:2601–2603.
84. Watanabe, K., S. Takesue, and K. Ishibashi. 1977. Reversibility of the adsorption of bacteriophage PL-1 to the cell walls isolated from *Lactobacillus casei*. J. Gen. Virol. 34:189–194.
85. Yasukawa, H., M. Kakikawa, Y. Masamune, A. Taketo, and K. I. Kodaira. 1997. Purification and DNA-binding properties of the integrase protein Int encoded by *Lactobacillus plantarum* phage ϕg1e. Gene 204:219–225.
86. Yoon, S. S., R. Barrangou-Poueys, F. Breidt, Jr., T. R. Klaenhammer, and H. P. Fleming. 2002. Isolation and characterization of bacteriophages from fermenting sauerkraut. Appl. Environ. Microbiol. 68:973–976.
87. Yoon, S. S., J. W. Kim, F. Breidt, and H. P. Fleming. 2001. Characterization of a lytic *Lactobacillus plantarum* bacteriophage and molecular cloning of a lysin gene in *Escherichia coli*. Int. J. Food Microbiol. 65:63–74.

This page intentionally left blank

PART VI

APPLICATIONS

This page intentionally left blank

42

Control of Bacteriophage in Commercial Microbiology and Fermentation Facilities

GREGG BOGOSIAN

There are an estimated 5×10^{30} bacterial cells on Earth (29), suggesting a global population of about 1×10^{31} bacteriophages. Roger Hendrix, in his characteristic whimsical fashion, has calculated that if bacteriophages were the size of cockroaches, they would cover the surface of the Earth in a layer 50,000 km deep. This type of thinking sends chills through those who work in commercial fermentation facilities, where bacteriophages really are considered cockroaches.

Experienced members of the commercial fermentation industry live in constant fear of bacteriophage contamination. Products obtained from the large-scale fermentation of bacteria include a wide variety of foods, pharmaceuticals, vitamins, solvents, enzymes, and more. The loss of an entire fermentation batch to bacteriophage contamination is a dramatic and unsettling event, often leading to such heavy bacteriophage contamination of the fermentation facility that further fermentation runs are not possible for an extended period of time. In addition, raw materials, isolation and purification equipment, and finished product may be contaminated by bacteriophage. The economic setbacks from product loss, raw material spoilage, and nonproductive operation costs can be substantial. Bacteriophage contamination can become an extremely distressing and frustrating problem for fermentation-based industries. There have been instances when bacteriophage outbreaks have taken months to bring under control. This chapter explores the possible causes of bacteriophage outbreaks, and the most effective approaches to handling the problem.

Extent of the Problem of Bacteriophage in Industrial Fermentations

Industries based on large-scale bacterial fermentations can be sorted into three classes, representing increasing levels of culture purity and media sterility: food production, commodity chemical production, and the biotechnology and pharmaceutical industries. Materials such as silage, compost, and treated sewage can also be considered products of large-scale bacterial fermentation, and are subject to upset by bacteriophage infestation, but the scope of this chapter is the threat of bacteriophage to the growth of commercial bacterial strains in controlled fermentation vessels.

In principle, any bacterial population is susceptible to attack by bacteriophage, and all three classes of the commercial fermentation industry suffer from the problem. By far the greatest problem is in the food industry, especially in the production of cheeses, yogurt and other fermented dairy products by lactic acid bacteria (mostly streptococci). The number of dairy-product fermentation batches lost to bacteriophage has been estimated to be in the range of 1–10%, probably due to the nonsterile fermentation conditions employed, along with raw starting materials that can be contaminated with bacteriophage at levels too low to be detected. The industries employing bacterial fermentations to produce commodity chemicals, including amino acids, vitamins, organic acids and alcohols, enzymes, solvents, biopolymers, and antibiotics, have also experienced considerable problems with bacteriophage. The struggles with bacteriophage contamination and control in these older fermentation industries have been extensively reviewed (1, 4, 5, 10, 11, 16, 17, 19, 21–23).

The advent of the recombinant DNA era in the late 1970s gave rise to the new biotechnology industry. By the early 1980s this industry was manufacturing commercial products from fermentation of recombinant bacteria, especially *Escherichia coli* strains. There has been only general mention in two of the more recent reviews (19, 23) of problems with bacteriophage in the biotechnology industry for three reasons. First, the industry is only about 20 years old. Second, the high containment employed for these fermenters provides a higher degree of protection from bacteriophage contamination. Third, no papers on bacteriophage problems in the biotechnology industry have been published to date.

For this review the author drew on his own experiences and those of his colleagues from 20 years in the trenches with bacteriophage, and also (with the generous help of the Society for Industrial Microbiology) queried many individuals in the biotechnology fermentation industry about encounters with bacteriophage. Representatives of 34 companies replied to the query, the results of this survey being that 24 companies reported they had experienced fermentation losses from bacteriophage and 10 reported they had not. The number of biotechnology fermentation batches lost to bacteriophage was about 0.1%, which is 1–2 orders of magnitude less than in the dairy-products fermentation industry. Many of the respondents provided unpublished observations that have been incorporated into this chapter.

Symptoms of a Bacteriophage-Contaminated Fermentation Process

In a controlled fermentation vessel, there are many possible symptoms of attack by bacteriophage. These symptoms, of which several may be present, are due to the inhibition of bacterial growth and lysis of the culture. They include a drop in turbidity, an increase in viscosity, excessive foaming, a rise in dissolved oxygen, a drop in carbon dioxide generation, and decreased demand for temperature control, pH adjustment, and nutrients such as ammonia or the carbon source.

Trained and experienced fermentation personnel can recognize bacteriophage contamination by both direct visual inspection and ordinary light microscopic examination of a sample of the culture. The sample may be more viscous than normal, and may appear grainy. In some cases, the turbidity of the culture may be so reduced that the sample takes on the appearance of uninoculated culture medium with particulate matter in it. Microscopic examination may reveal that the sample contains fewer than the normal number of bacterial cells, and cell debris may be present. Prolonged movement (streaming) of the cells and debris may be observed, behavior associated with the high viscosity of the sample.

Further confirmation of the presence of bacteriophage can be obtained in an easy but not necessarily quick test. The suspect culture is centrifuged at low speed to pellet the cells and debris, and passed through a 0.2 μm filter to ensure complete removal of cells. A small amount of filtrate (e.g., one drop, or 50 μl) is then spotted on a lawn of appropriate indicator bacterial cells, usually prepared in a top agar. To ensure appropriate susceptibility to any bacteriophage present, the best choice as bacterial indicator would be the same strain that is being fermented. One or more plaques in the area of the spotted filtrate indicates the presence of bacteriophage, but this is usually not evident until after overnight incubation. If the filtrate contains many bacteriophage, then the spotted area will appear as a clear area in the lawn due to the presence of overlapping plaques. It is possible with lawns that have been prepared hours ahead of time to get results in as little as 1 hour. The test can also be run as a mixed pour plate by mixing a small volume (e.g., 0.1 ml) of the filtrate with the bacterial cells and top agar before pouring the lawn. In this case, a positive test would be the appearance of one or more plaques in the lawn, or many overlapping plaques and thus confluent lysis and total clearing of the plate.

The appearance of individual plaques can be taken as a definitive finding of bacteriophage. However, it should be kept in mind that other fermentation upsets, such as the inadvertent loss of pH control or the introduction of an inhibitory chemical such as a detergent or an antibiotic, can give similar fermentation symptoms and local clearing of the bacterial lawn in a spot test or total clearing in a mixed pour plate test. If there is any reason to suspect such a problem, a more rigorous bacteriophage test would be to prepare several 10-fold dilutions of the culture filtrate, and mix these with the indicator bacterial cells and the top agar before pouring the lawn. In such a test, true bacteriophage contamination would be evident by the presence of individual plaques at the higher dilutions.

Consequences of Bacteriophage Contamination

Low levels of bacteriophage contamination do not precipitate the fermentation irregularities listed above, and may escape notice in the absence of a spot test. Indeed, in the food and some of the commodity chemical fermentation industries, low-level bacteriophage contamination is tolerable in that acceptable product is still obtained. For almost all of the biotechnology industry, however, no level of bacteriophage contamination is acceptable due to the potential for process and product adulteration, and every fermentation batch is subjected to a bacteriophage test as a standard quality control measure.

Fermentation batches lost to bacteriophage mean the loss of both product and raw materials used to prepare the culture medium. In addition, there are the costs associated with the cleanup and the losses resulting from any downtime in the facility. Subsequent quality control investigations and implementation of corrective action plans add to the costs associated with a bacteriophage outbreak. With a valuable fermentation product, such as a pharmaceutical, the cost of a brief encounter with bacteriophage could be in the millions of dollars.

It is reassuring to inform facility personnel that bacteriophage pose absolutely no risk to humans. Indeed, there is renewed interest in bacteriophage therapy and other applications of bacteriophage, where studies have shown that exposure to bacteriophage does not pose any risk to humans (3, 12, 26, 27; and chapter 48).

Control of Bacteriophage Contamination

Keeping Bacteriophage Out of a Fermentation Facility

Most fermenters are designed to be closed systems and the contents would not normally be exposed to environmental contaminants such as bacteriophage. However, no system is perfectly closed, and several potential sources of bacteriophage entry have been identified.

The culture medium in the fermenter is normally heat-sterilized in place, and would be bacteriophage-free. The addition to the fermenter of nonsterile solutions, or solutions which have been filter-sterilized, could introduce bacteriophage. The inocula should not be overlooked in this respect, and the cell banks used to prepare the inocula should be tested for bacteriophage before use. Some unsterilized raw materials are essentially self-sterilizing; for example, ammonium hydroxide at the standard concentration of 29%, commonly used for pH control and as a nitrogen supplement, inactivates bacteriophage with a decimal reduction time (the time required for a 10-fold drop in bacteriophage titer) of about 2 seconds. Solutions with a pH of less than 5 or greater than 10 would have similar self-sterilizing behavior with respect to most bacteriophage (18).

For many biotechnology fermentation culture media, antibiotics are included to maintain selective pressure for the plasmid expression system. These antibiotics are also very effective at keeping the culture free of bacterial contaminants, so much so that complacency may set in among the operators. It is imperative that facility personnel be reminded not to relax their vigilance against bacteriophage, which of course would be completely impervious to antibiotics.

The fermenter components should be maintained in good working order and inspected regularly. One common source of bacteriophage is chilled water, running through coils in the fermenter to control culture temperature; leaking cooling coils could introduce bacteriophage. In some facilities, the frequency of bacteriophage attack correlated with bacterial load in the chilled water. The bacteria were not necessarily directly related to the bacteriophage but, rather, served as an indicator of the effectiveness of the water-treatment system supplying the chilled water. The occurrence of an attack was also correlated with the chilled water having a higher than normal biological oxygen demand (BOD), indicating a high organic matter content. Additional water treatment, maintenance, and purging of the system eliminated the problem. Sodium hypochlorite is sometimes added at very low concentrations to chilled water systems to control microbial contamination, but the corrosiveness of this chemical may cause leaks in the cooling coils.

Potential bacteriophage contamination of the air supply is a critical weak point, being difficult to treat. Air intake filters, even though their average pore size is slightly larger than most bacteriophage, can remove low levels of bacteriophage from the incoming air (21). The location of the air intake system is critical in minimizing the numbers of bacteriophage entering the system (see below). Humid air (or wet weather) makes air filters less effective. It has also been observed that if air filters are depressurized too fast after sterilization, then they can rupture and be rendered much less effective. Frequent integrity testing of air filters is also advisable. Heat treatment of incoming air is highly effective against bacteriophage, but not practical when large air volumes are required. Good microbiological and cleaning practices (see below for cleaning recommendations) help to keep bacteriophage out of the fermentation facility. While it may seem obvious to employ cleaning agents that include a disinfectant effective against bacteriophage, agents which are effective against bacteria but not against bacteriophage (see below) still have utility. It is important to prevent host bacteria from contaminating the facility and providing a reservoir where bacteriophage could propagate.

Monitoring within the facility for the presence of bacteriophage should be performed at least once every 2 weeks. In this procedure, small pieces of sterile absorbent filter paper are used to wipe the test area, then are soaked in water to extract any bacteriophage that may have been picked up, the extract passed through a 0.2 μm filter to remove larger particles, and the filtrate used in a spot test or mixed pour plate test. These tests are described above. The same criteria should be used to interpret the test results, including consideration of the possibility of false positive results from the wipe test picking up a chemical inhibitory to the indicator bacteria. Any positive results should provoke a bacteriophage action plan, consisting at least of increased cleaning and monitoring. Key areas to monitor and clean include floors, sinks, drains, and any instruments or areas where the bacterial culture is handled or samples stored, such as spectrophotometers, centrifuges, and refrigerators.

The design of the fermentation facility is important in minimizing the risk of a bacteriophage outbreak. It should be easy to clean and keep dry. Most bacteriophage are susceptible to desiccation but can exist indefinitely in wet environments. Thus, standing water and damp areas should be eliminated. Spills should be cleaned up, and leaks fixed promptly. Hidden wet areas should be identified and dealt with. Floor mats are often overlooked as areas that can retain dampness. The floor surface should be of a durable material which resists cracking and pitting, and which is easy to keep spotlessly clean and dry.

In one research facility for *E. coli* fermentations, a holding tank in the waste system contributed to a chronic bacteriophage problem. All of the fermentation wastes, including live *E. coli* cells, entered the holding tank. When the tank became full, the contents were pumped out, sterilized by passage through a heat-exchanger, and then directly dumped into a sewer system. Thus, the tank itself was never sterilized. This holding tank supported a resident bacteriophage population of about 1×10^5 plaque forming units (PFU) per milliliter, presumably replicating at the expense of the live *E. coli* cells regularly entering the tank. The bacteriophage would occasionally find their way up the drains and into the working areas of the facility. Interestingly, the resident bacteriophage were initially of one type, but less than 1 year later had been replaced by a completely different bacteriophage, indicating the occurrence of species succession. This second bacteriophage appeared to adapt to the facility in that within a year a variant form appeared which had a spontaneous deletion of about 1 kb from the genome of about 49 kb. This new form of the second bacteriophage immediately replaced the old form, and remained resident in the facility for several years. The solution to this problem was a redesign in which the tank could be drained and then sprayed internally with a hot caustic disinfectant (see below). Obviously, it is very important to eliminate any possibility of chronic bacteriophage contamination that can lead to species succession and adaptation to the facility.

There has been considerable work on the utilization of fermentation medium components that might inactivate or inhibit the replication of any bacteriophage that enter the fermenter (1, 10, 11, 16, 30). The approaches include attempting to inhibit bacteriophage absorption to bacterial cells by using low-calcium and low-magnesium media; metal ion chelators such as citrate, oxalate, or phytic acid; nonionic detergents such as PEG, Tween-20, and Tween-80; and adding inhibitors of bacteriophage nucleic acid injection such as sodium tripolyphosphate. To date, these approaches have been limited by the components being expensive, inefficient, or overly specific.

Utilization of Bacteriophage-Resistant Strains

The use of bacteriophage-resistant strains seems an obvious approach to bacteriophage control. However, there are several drawbacks. Strains resistant to multiple classes of bacteriophage are frequently harder or impossible to manipulate by genetic transduction. Resistant strains have also been observed to have lower product yields and altered flow characteristics affecting downstream processing. The problem bacteriophage also can adapt to the initially resistant host, rendering it useless.

There are three approaches to effectively and safely using bacteriophage-resistant strains. One is to hold them in reserve, only to be used during a bacteriophage crisis. A second is to use a mixed culture of strains, each resistant to a different class of bacteriophage, in conjunction with a rotation schedule in the event of a bacteriophage outbreak (24, 25, 28). The third is to use strains with well-understood resistance mechanisms (1) that bacteriophage would not be able to adapt to. For example, it is a widespread practice in the biotechnology industry to use strains of *E. coli* with nonrevertible mutations in *fhuA* (the outer membrane receptor for T-odd bacteriophage) in order to protect against the dreaded bacteriophage T1 (see below and chapters 17, 19, and 20 which cover the biology of phages T1, T5, and T3/T7, respectively).

Treating a Fermenter Which Has Been Lost to Bacteriophage

While some bacteriophage are relatively heat-resistant, all are rapidly inactivated in solution when heated to over 80°C (18), and thus a contaminated fermenter culture should be heat-sterilized in place. Hot caustic solutions (for example, 0.1 M sodium hydroxide at 50°C) are effective at eliminating bacteriophage from tanks and other equipment that can withstand such treatment. Apparently all commonly encountered bacteriophage are extremely sensitive to acidic (pH 4 or lower) or alkaline (pH 11 or higher) conditions (18). While it is thus a fairly straightforward matter to inactivate all of the bacteriophage inside a contaminated fermenter, with bacteriophage titers in fermenters as high as 1×10^{12} PFU/ml the facility around the affected fermenter typically is also heavily contaminated. Such contamination of the facility usually occurs as a result of routine sampling of the fermenter culture before the bacteriophage attack is recognized, as well as from any leaks that may be present. During one outbreak, a water puddle underneath a contaminated fermenter was found to contain bacteriophage at about 1×10^8 PFU/ml. The fermenter exhaust system can also spread large numbers of bacteriophage, making it prudent to implement methods to contain the exhaust gases. Such measures would include shutting off the airflow as soon as possible, watching for and disinfecting any foam that exits the fermenter, and using either a scrubber or a filter system on the exhaust system. The next section discusses strategies for cleaning and disinfecting the rest of the facility.

Disinfection of Bacteriophage-Contaminated Facilities

A bacteriophage-contaminated facility requires extensive and repeated cleaning. The cleaning agents employed must include disinfectants effective against bacteriophage. Wet and damp areas deserve special attention, but everything in the facility should be cleaned or disposed of if possible. Mopping, swabbing, and wiping with generous quantities of disinfectant solutions is the general approach.

There have been numerous studies on the effects of various agents on bacteriophage, and several reviews have compiled lists of chemicals, radiation, etc., to which bacteriophage are susceptible (1, 2, 9, 14, 20, 21). However, most of the listed agents are limited by being too toxic, too expensive, too specific, or only partially effective. Most notably, ethanol is relatively ineffective against most bacteriophage, with decimal reduction times (the time required for a 10-fold drop in bacteriophage titer) for 100% ethanol of about 2 days and for 70% ethanol of about 12 hours. Germicidal UV light is highly effective, but only for surfaces it can reach, imparting limited utility. Widely used disinfectants with the active ingredients sodium *o*-phenylphenate and sodium *p*-tertiary amylphenate (such as the commercial disinfectant Vesphene II) are totally ineffective against bacteriophage, but do have utility in eliminating bacterial reservoirs of bacteriophage propagation.

The bacterial fermentation industry needs bacteriophage disinfectants that are relatively safe for the human operators, inexpensive, and highly effective against a broad range of bacteriophage. There are three types of disinfectants that fit the bill. Sodium hypochlorite at 0.05% (for example, a 1:100 dilution of commercial bleaches such as Clorox) and formaldehyde (formalin) at 0.02% (for example, a 1:100 dilution of the commonly used commercial disinfectant DC&R) are both highly effective against apparently all types of bacteriophage and are widely used for this purpose. These agents have decimal reduction times against bacteriophage of about 10 seconds. Both of these agents are also highly effective antibacterials. Drawbacks of these two agents are the relative toxicity of formaldehyde and the relative corrosiveness of sodium hypochlorite.

The third good bacteriophage disinfectant is less well known, but should be more widely employed. It is ascorbic acid (vitamin C) with trace amounts of copper to catalyze auto-oxidation (13, 15). We call it “ascorbinator” because it is so effective against bacteriophage. The working solution is 10 mM ascorbic acid (1.76 g/l) and 0.05 mM cupric chloride (1.7 mg/l); the pH as prepared is about 3.2, and is not adjusted further. Cupric sulfate can be substituted for the cupric chloride. The solution can be stored at room temperature for up to 1 week after being prepared. Ascorbinator is highly effective against apparently all types of bacteriophage, and exhibits a decimal reduction time against bacteriophage of about 4 seconds. It retains this level of effectiveness even when diluted 10-fold, and thus can be used to disinfect wet areas. Raising the pH above 5, however, increases the decimal reduction time to about 2.5 minutes. Ascorbinator is relatively ineffective as an antibacterial. It is very safe for the human operators, and is not corrosive like sodium hypochlorite.

The mechanism of action of ascorbinator appears to be the oxygen-dependent formation of free radicals during the auto-oxidation of ascorbic acid, which cause single-strand scissions in nucleic acids (13). It has long been known that ascorbic acid is effective against bacteriophage (2) and that copper enhances the effectiveness of ascorbic acid as an antimicrobial (31); the subject is reviewed in more detail by Eller et al.(6).

During clean-up efforts, movement of personnel within the facility should be controlled, and nonessential personnel kept out. It has frequently been observed that shoes are a major route of bacteriophage spread; shoe covers and changing areas are effective at minimizing this spread, as are shallow stepping trays or mats filled with a disinfectant. The movement of wheeled carts, forklifts, and the like should also be controlled or steps taken to disinfect their wheels.

It is not possible by cleaning alone to completely eliminate every bacteriophage particle from a contaminated facility. Cleaning efforts should be as thorough as possible, with the recognition that those bacteriophage that are missed will become inactivated over time. Dried on a surface, bacteriophage T7 has a decimal reduction time of about 1 hour (18). Most bacteriophage on dry surfaces have decimal reduction times ranging from 1 hour to 10 days, and thus will eventually disappear from a contaminated facility. The one legendary exception is bacteriophage T1, which when dry has been reported to persist almost indefinitely (18). Only at high temperatures is appreciable inactivation of T1 observed, with a decimal reduction time on a dry surface at 90°C of about 14 hours (18). As noted above, the main line of defense against a bacteriophage T1 outbreak is to employ resistant bacterial strains.

Factors that Increase the Risk of a Bacteriophage Outbreak

The external environments of soil, water, and sewage are a rich source of bacteriophage (8; and chapters 5, 33, and 45) and should be kept out of the fermentation facility. Since bacteriophage are dependent on bacteria for growth and reproduction, good control of the strain being fermented will help limit populations of bacteriophage which can attack it. Good sample handling and storage, containment, and waste handling practices are essential. At one facility, a bacteriophage outbreak occurred every time the fermenter exhaust filter was not changed on schedule, or was inadvertently not installed at all. Apparently, the release of bacterial cells in the exhaust contributed to the propagation of high bacteriophage populations in the environment near the air intake. A similar situation was found in another facility where the air intakes were downwind of the bacterial cell processing plant. For the recombinant *E. coli* strains used in the biotechnology industry, sewage provides a dangerous source of bacteriophage

(7, 8, and chapter 45). One *E. coli* facility, with air intakes located downwind from a sewage treatment plant, suffered frequent bacteriophage outbreaks. Another *E. coli* facility had air intakes located near a dairy manure lagoon, with coliphage present in the lagoon at about 200 PFU/ml, and consequent bacteriophage outbreaks until the air intake was relocated.

There is a seasonal aspect to bacteriophage outbreaks, with peaks occurring in January–March, and again in October–November. One *E. coli* facility that suffered five bacteriophage outbreaks in 14 years had one in January, one in March, two in October, and one in November (in different years in each case). It has been noted that there are increased numbers of bacteriophage in soil and sewage during these times of the year (16). Tom Anderson has suggested that spring planting and fall harvesting activities may release bacteriophage from disrupted soils. Sherwood Casjens has suggested that the migrations of birds may play a role.

Disruption of soil in any season may contribute to bacteriophage outbreaks. Several facilities have suffered outbreaks during ground-breaking for construction near the fermentation areas. In one such case a sewer line was accidentally broken during the construction work, an event that was quickly followed by a bacteriophage outbreak in the facility. In another case, at an *E. coli* facility suffering an outbreak, the culprit coliphage were found at a level of about 300 PFU/ml in a puddle at a nearby construction site (although it was not possible to pinpoint which was the cause and which the effect). One control measure to consider would be to cover disturbed ground at construction sites in order to minimize the release of bacteriophage.

Salvage of Bacteriophage-Contaminated Cultures

For most of the biotechnology industry, bacteriophage contamination dictates that the culture be discarded and any potentially contaminated product destroyed. For some food and commodity chemical products, if the culture viscosity or residual nutrient levels are not too high to affect downstream processing, and if the product titer is high enough to economically process, then the batch can be salvaged.

There has been limited inquiry into the possibility of adding chemicals to a bacteriophage-contaminated culture which would not significantly affect the bacterial strain but which would destroy the bacteriophage or interfere with their replication (1, 30). Such approaches would require either a rapid on-line test for the presence of bacteriophage, or a fermentation process that is limping along with a low-level chronic bacteriophage contamination. It is unlikely that any such approach would be acceptable in the biotechnology industry.

Concluding Remarks

Fermentation facility personnel and laboratory staff should be trained to expect and respond to bacteriophage outbreaks. While the problem may seem unavoidable in principle, some facilities have run bacteriophage-free for several years. They practice effective cleaning and monitoring, and do not relax their guard or become complacent. It is advisable to increase cleaning and monitoring during "bacteriophage season", January–March and October–November, and during any significant ground-breaking near the facility. Prevention measures and response plans to a positive monitoring result or the loss of a fermenter vessel to bacteriophage contamination should be in the form of written standard operating procedures. Eternal vigilance is the price of freedom from bacteriophage.

Acknowledgments

Many people contributed unpublished observations to this chapter, or helped identify others who could provide such information. For these contributions I thank Tom Anderson, Carl Bruch, Harald Brüssow, Rich Calendar, Sherwood Casjens, Jennie Hunter Cevera, Shiau-Ta Chung, Paul Cino, Bill Cover, Roy Curtiss, Jac Dekker, David Dodds, Donna Duckworth, Craig Gershater, Hugh Graham, Rowan Grayling, Patricia Hanretta, Roger Hendrix, Osama Ibrahim, Lonnie Ingram, Jim Kane, Todd Klaenhammer, Jeanette Leach, John Mott, Tadhg O'Sullivan, Earl Pursell, Mary Ellen Sanders, Bill Sandine, Rob Saunders, Ron Somerville, Jaroslav Spizek, Ray Testa, Joe Tunac, Karen Van Note, Goutham Vemuri, and Charley Yanofsky. I thank Gary Bond, Greg Gibb, Patti Morris, Julia O'Neil, and Todd Staley for critically reviewing the manuscript.

References

1. Ackermann, H.-W., and M. S. DuBow. 1987. Viruses of Prokaryotes. CRC Press, Boca Raton, Fla.
2. Adams, M. H. 1959. Bacteriophages. Interscience Publishers, New York.
3. Barrow, P. A., and J. S. Soothill. 1997. Bacteriophage therapy and prophylaxis: rediscovery and renewed assessment of potential. Trends Microbiol. 5:268–271.
4. Bradley, S. G., and L. A. Jones. 1968 Bacteriophages, their biology and industrial significance. Prog. Ind. Microbiol. 7: 43–75.
5. Brussow, H. 2001. Phages of dairy bacteria. Annu. Rev. Microbiol. 55:283–303.
6. Eller, C., F. F. Edwards, and E. S. Wynne. 1968. Sporicidal action of auto-oxidized ascorbic acid for *Clostridium*. Appl. Microbiol. 16:349–354.
7. Ewert, D. L., and M. J. B. Paynter. 1980. Enumeration of bacteriophages and host bacteria in sewage and the

activated-sludge treatment process. Appl. Environ. Microbiol. 39:576–583.

8. Goyal, S. M., C. P. Gerba, and G. Bitton (eds.). 1987. Phage Ecology. Wiley, New York.
9. Hall, E. A., F. Kavanagh, and I. N. Asheshov. 1951. Action of forty-five antibacterial substances on bacterial viruses. Antibiot.Chemother. 1:369–378.
10. Hongo, M. 1971. Phage contamination and control in the fermentation industry, pp. 233–265. *In* K. Sakaguchi, T. Uemura, and S. Kinoshita (eds.) Biochemical and Industrial Aspects of Fermentation. Kodansha, Tokyo.
11. Hongo, M. 1971. Phage contamination and control, pp. 67–90. *In* K. Yamada (ed.) Microbial Production of Amino Acids. Kodansha, Tokyo.
12. Levin, B. R., and J. J. Bull. 1996. Phage therapy revisited: the population biology of a bacterial infection and its treatment with bacteriophage and antibiotics. Am.Nat. 147:881–898.
13. Murata, A., K. Kitagawa, and R. Saruno. 1971. Inactivation of bacteriophages by ascorbic acid. Agr. Biol. Chem. 36:1065–1067.
14. Murata, A., K. Kitagawa, and R. Saruno. 1972. Inactivation of bacteriophages by thiol reducing agents. Agr. Biol. Chem. 35:294–296.
15. Murata, A., and K. Kitagawa. 1973. Mechanism of inactivation of bacteriophage J1 by ascorbic acid. Agr. Biol. Chem. 37:1145–1151.
16. Ogata, S. 1980. Bacteriophage contamination in industrial processes. Biotechnol. Bioeng. 22 (Suppl. 1):177–193.
17. Peitersen, N. 1991. Practical phage control. Bull. Int. Dairy Fed. 263:1–43.
18. Pollard, E., and M. Reaume. 1951. Thermal inactivation of bacterial viruses. Arch. Biochem. 32:278–287.
19. Primrose, S. B. 1990. Controlling bacteriophage infections in industrial bioprocesses, pp. 1–10. *In* J. Reiser (ed.) Applied Molecular Genetics. Springer, Berlin.
20. Putman, F. W. 1953. Bacteriophages: nature and reproduction, pp. 177–284. *In* M. L. Anson, K. Bailey, and J. T. Edsall (eds.) Advances in Protein Biochemistry, vol. 8. Academic Press, New York.
21. Rudolph, V. 1978. Bacteriophages in fermentation. Process Biochem. [Jan]:16–26.
22. Sanders, M. E. 1987. Bacteriophages of industrial importance, pp. 211–214. *In* S. M. Goyal, C. P. Gerba, and G. Bitton (eds.) Phage Ecology. Wiley, New York.
23. Sanders, M. E. 1999. Bacteriophages in industrial fermentations, pp. 116–121. *In* H. Frankel-Conrat and P .C. Kimball (eds.) Encyclopedia of Virology. Harcourt Brace, New York.
24. Sandine, W. E. 1985. Cheese 7, viruses 0. Sci. Food Agric. 3:27–30.
25. Sandine, W. E. 1989. Use of bacteriophage-resistant mutants of lactococcal starters in cheesemaking. Neth. Milk Dairy J. 43:211–219.
26. Sharp, R. J., and P. Hambleton. 2001. A coming of phage: bacteriophages in biotechnology. J. Chem. Technol. Biotechnol. 76:661–699.
27. Summers, W. C. 2001. Bacteriophage therapy. Annu. Rev. Microbiol. 55:437–451.
28. Thunell, R. K., W. E. Sandine, and F. W. Bodyfelt. 1981. Phage-insensitive, multiple-strain starter approach to cheddar cheese making. J. Dairy Sci. 64:2270–2277.
29. Whitman, W. B., D. C. Coleman, and W. J. Wiebe. 1998. Prokaryotes: the unseen majority. Proc. Natl. Acad. Sci. USA 95:6578–6583.
30. Yamanaka, S., N. Kashima, and K. Mitsugi. 1975. Method for limiting damage due to bacteriophages in fermentation media. United States Patent 3,880,718.
31. Zimmerman, L. 1966. Toxicity of copper and ascorbic acid to *Serratia marcescens*. J. Bacteriol. 91:1537–1542.

43

Phage-Based Expression Systems

NOREEN E. MURRAY

Phage, and plasmid, derivatives that had "picked up" genes from the *Escherichia coli* chromosome led Campbell (12) to propose his influential model based on recombination between circular genomes (see chapter 7). According to this model, segments of bacterial DNA were added to a phage genome if excision of the prophage was by an "aberrant" recombination event. These unexpected, or *illegitimate* (2, 27, 96) events fortuitously created the early recombinant clones that became tools at the cutting edge of research in the pioneering days of molecular biology. They, for example, facilitated the analyses of bacterial operons (30, 61) and enabled the amplification of the Lac repressor (60). Monitoring the bacterial enzymes specified by λ*trp* phages enhanced our understanding of the regulatory mechanisms of bacteriophage λ (18, 28, 29), telling us much about the control of transcription from the early λ promoters and suggesting how a phage promoter might be harnessed to amplify gene products. While these first "recombinant clones" of phages and plasmids were of unpredictable content, by 1974 it was possible to design and make new combinations of genes from any source in either plasmid (15) or phage λ vectors (62, 76, 92). This article will introduce the use of λ promoters within the context of the phage genome and follow the use of these and other phage promoters as components of genetically engineered plasmids, where they are relieved of their essential roles in the normal life cycle of the phage and instead used to amplify gene products.

Expression of "Foreign" Genes in λ Phages

Phage λ Itself

Genes cloned in phage λ, whether via in vivo or in vitro recombination, are commonly positioned within the "central region" of the genome between genes *J* and *N* (see figure 43-1), where the DNA may be transcribed soon after infection from the leftward promoter (p_L) or, much later, from the rightward promoter (p_R') (see chapter 27). Transcription from the early promoters of phage λ (p_L and p_R) is susceptible to negative control by two phage proteins: repressor (CI) and Cro (see chapter 8). While CI is essential only for lysogeny, Cro protein moderates transcription from p_L and p_R during the lytic pathway (25, 28, 70). Failure to moderate the transcription of early genes has a detrimental effect on the lytic pathway, including DNA replication (26). During the normal lytic pathway, concomitant with the moderation of transcription from early promoters, Q protein serves to allow transcription initiated at p_R' to proceed beyond a transcription-terminator signal (figure 43-1) that separates p_R' from the late genes (99). Q protein is essential for transcription of the late genes as a single transcript that continues through the late genes into the central region (8, 10) (see chapter 9).

Early experiments demonstrated that transcription initiated at p_L can proceed through sequences that normally signal the termination of transcription (1, 29, 83). This robust transcription is dependent on the interaction of N protein, the product of the phage gene adjacent to p_L, with RNA polymerase at a specific sequence (*nut*$_L$) within the transcript (81). Q protein also serves to enable RNA polymerase to read through termination signals (99). However, neither N- nor Q-modified RNA polymerase is blind to all terminator signals. The central region of the phage genome includes sequences that are orientation-dependent terminators of transcription from p_L and p_R' (10, 33, 37, 45) (see figure 43-1). Effective transcription from these promoters therefore will be prevented if a native terminator sequence is retained between the phage promoter and the inserted coding sequence of interest, but transcription is likely to be unimpeded by potential terminator sequences introduced from elsewhere. In early cloning experiments the relevant coding sequence was usually flanked by significant stretches of additional DNA thereby increasing the chance of interference from a transcription terminator.

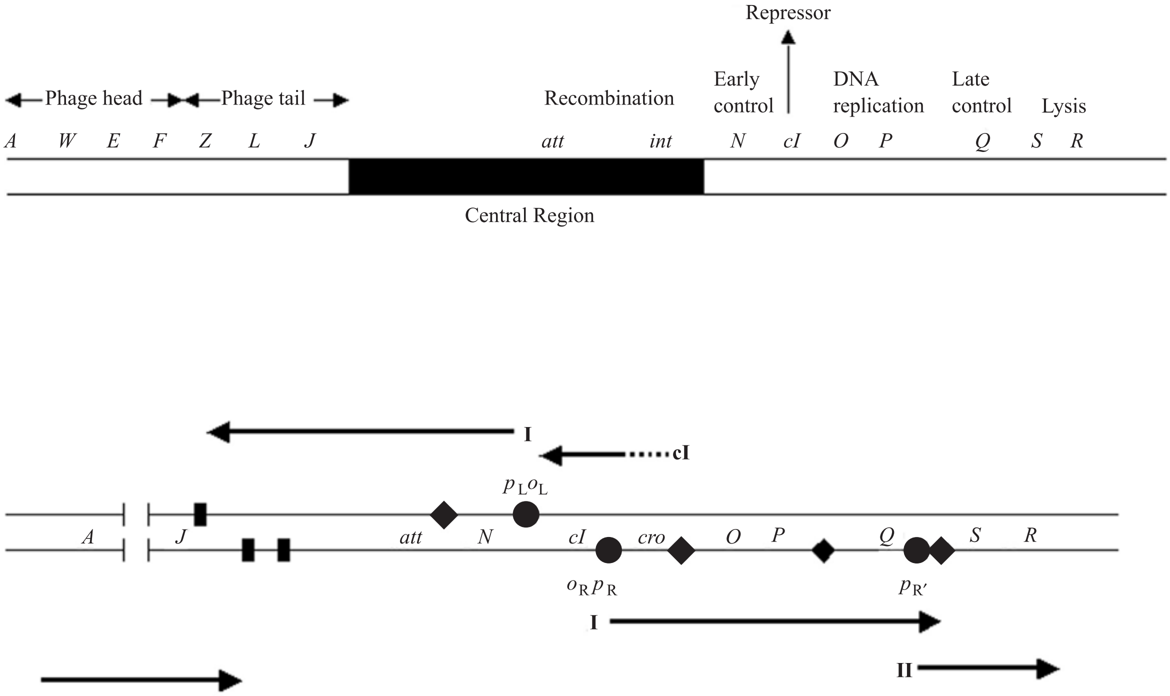


Figure 43-1 Phage λ: its genome and transcription circuits. Top: The phage DNA enters the bacterium as a linear molecule but circularizes via its cohesive ends. Genes that encode related functions are clustered in the λ genome. Genes within the "central region," identified by the black segment, are inessential for propagation of the phage. *att*, located within the central region, identifies the site at which site-specific recombination integrates the phage genome into the bacterial chromosome. The map is not drawn to scale (for a more detailed look at the phage λ genome, see figure 27-2.) Bottom: Transcription of the genome proceeds initially from two promoters, p_L and p_R. At the earliest stage of infection, transcripts initiated at p_L and p_R (I) terminate at specific signals, indicated by black diamonds ♦ (some rightward transcripts initiated at p_R reach the second termination signal). Only in the presence of the *N* gene product does most transcription proceed beyond the termination signals into the downstream genes. The consequent synthesis of Q protein allows transcription of the late genes (II) to proceed from p_R' through the adjacent termination signal into the lysis genes and the genes that specify the components of the phage head and tail. Transcription of the repressor gene (*cI*) during the establishment of lysogeny is initiated at a promoter, p_{RE}, situated to the right of *cro*. Repressor synthesis is maintained by autogenously controlled transcription from p_{RM}, a promoter that overlaps o_R. Indicated are major promoters (●); major termination signals susceptible to N- and/or Q-dependent antitermination (♦); and major termination signals that block transcription of N- and/or Q-modified RNA polymerase (■).

Gene Expression in λ

The use of p_L for amplifying gene expression could be favored by the role of N protein in rendering DNA polymerase insensitive to barriers that would otherwise impair transcription. Such insensitivity provides a high maximal rate of transcription, possibly an order of magnitude higher than that from the promoter of the *trp* operon of *E. coli* (18, 59). This maximal rate, however, is that obtained in the absence of both CI and Cro. In practice, the maximal exploitation of p_L is not readily reconciled with either Cro-moderated transcription during the normal life cycle of the phage or the impaired DNA replication seen in the absence of Cro (6, 64).

Many early experiments reporting the amplification of gene expression from the p_L promoter of phage λ depended on infection of bacteria with high-titer lysates of phage (43, 59). Infection with high-titer lysates can, in part, alleviate the need for good replication of the phage DNA. This may be useful for a Cro$^-$ phage, but a more experimenter-friendly route is via induction of a dormant prophage. The efficient excision of the prophage from the bacterial chromosome following induction requires that the phage retains its attachment site and associated site-specific recombination functions. Such derivatives of λ are readily made by the insertion of DNA fragments within the central region to the left of the attachment site (see figure 43-1), where in one orientation a foreign gene can be expressed by N-dependent transcription from p_L (38) and in the other orientation by Q-dependent transcription from p_R' (63).

Either infection of *E. coli* by phage λ in the absence of repressor, or induction of a prophage, normally leads to lysis of the cell within an hour. To maximize the yield of the product from a bacterial gene within λ, it is essential to delay lysis (61). This can be achieved by a mutation in gene *S*. A defect in this phage gene prevents lysis of the cell but permits DNA replication and allows protein synthesis to continue for hours (34) (see chapter 10). Amplification of gene products from p_L has used phages defective in both gene *S* and gene *Q* (38, 59, 64). In the absence of Q protein, transcription of the late genes, including gene *S*, is not activated, and the protein-synthesizing machinery of the cell is not diverted to the synthesis of phage proteins. Amplification from p_R' has depended on vectors defective in gene *S* and gene *E* (55, 63). The mutation in gene *E* serves to block synthesis of the major component of the phage capsid, and the packaging of phage genomes.

Transcriptional Interference

Efficient transcription from phage promoters raises the problem of transcriptional inhibition by convergent transcription, particularly when aided by proteins that allow RNA polymerase to ignore terminator sequences. The effect of convergent transcription on gene expression has been monitored by experiments using the p_L and p_{trp} promoters within phage λ (95); *trp* expression is impaired by transcription from p_L and vice versa. Many expression vectors include a transcription termination signal that prevents transcription from the strong promoter from either opposing or reinforcing transcription of other genes.

Amplification of Useful Enzymes

Examples in which induction of a λ prophage has been used to amplify useful enzymes include the DNA ligase of *E. coli* from its own promoter (67); DNA PolI from the combined use of its own promoter and p_L (64); and T4 DNA ligase (63) and T4 polynucleotide kinase (Pnk) (55) from p_R'. The genes for both T4 DNA ligase and Pnk are positioned close to the late genes, adjacent to gene *J*, and hence to the left of the natural terminators of rightwards transcription (figure 43-1). Efficient expression of *pol*A from p_R' was not obtained; the *pol*A gene in this phage is more remote from the late genes than those specifying T4 DNA ligase and Pnk from p_R' (64). It is possible that a terminator for rightward transcription was retained in the λ*pol* phage. The lysogenic strains referred to remain in commercial use, and provide levels of amplification approaching, or exceeding, 5% of soluble cell protein without any manipulation of the cloned DNA fragments.

Stability and Toxicity

The amplification of some proteins is toxic to bacterial cells. While the DNA polymerase I and DNA ligase genes were readily cloned in phage λ, and maintained within a prophage or during lytic infection, experiments indicated that their maintenance in a multicopy plasmid would need tight control of transcription. As predicted (43, 64, 77), the transfer of the functional *E. coli pol*A gene to a multicopy plasmid required the dissection of the gene from its own promoter (57).

Cloning the gene for T4 Pnk (*pse*T) illustrated that both position and orientation of foreign coding sequences with respect to the phage promoters can influence vegetative propagation of recombinant phages, and the recovery of foreign genes. The *pse*T gene was initially cloned following insertion of a relevant T4 DNA fragment within the *c*I gene of a λ vector in the opposite orientation to that in which *c*I is transcribed (figure 43-1); no rightward transcription of the *c*I region from a λ promoter is anticipated, and transcription of *pse*T will be opposed by transcription from p_{RE} (see legend to figure 43-1). The λ *pse*T phages were readily propagated, but phage growth was impaired when the DNA fragment containing *pse*T was transferred to the central region of the λ genome. Moreover, each subclone that was examined carried *pse*T in the orientation that avoids transcription from the early λ promoter, p_L (figure 43-1). Even so, the recombinant phages produced tiny plaques. In contrast, strains lysogenic for the recombinant phage grew well, and they were stable (55). The maintenance of a cloned gene as a single copy is a virtue of phage λ, a virtue that not only increases the stability of the cloned DNA but also facilitates genetic analysis of the coding sequence. Some new plasmid vectors (described later) now mimic λ in this respect, that is they are maintained in single copy, but their design permits amplification to a copy number of around 30 (84).

Expression Plasmids that Use a λ Promoter

Promoter p_L

Each of the three major promoters of phage λ has been used to drive transcription of a foreign gene within a plasmid vector (see figure 43-2 for examples). Transcription from p_L is regulated by CI, but plasmid vectors that use p_L usually retain *nut*$_L$, the site that permits N protein to modify RNA polymerase (see pKC30 in figure 43-2). Traditionally, the control of transcription, as with phage λ vectors, is mediated by a thermosensitive form of CI, rather than wild-type repressor, and relief of repression is achieved by raising the temperature from 32°C to 42°C. Host strains lysogenic for a defective (Cro$^-$) prophage provide the temperature-sensitive CI, and in one strain the defective prophage supplies N protein in the absence of Cro following the inactivation of repressor (5). The early *cro*$^-$ strains (13) included a deletion that extends from the prophage into the bacterial genome, leading to loss of the *uvrB* gene.

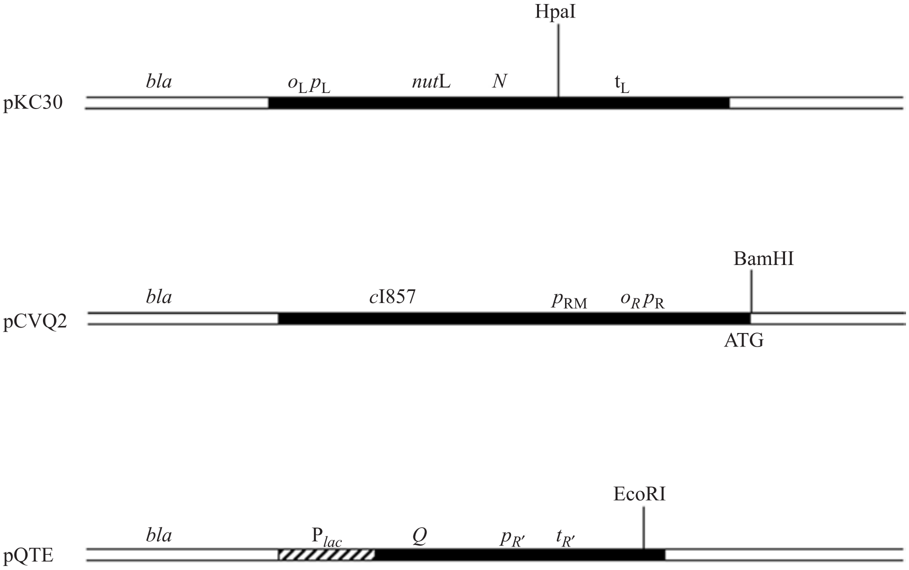


Figure 43-2 Simple examples of plasmid expression vectors using λ promoters. In each diagram the relevant elements of λ within the plasmid vector are located within the black segment. Top: Transcription in pKC30 is driven from p_L and is blocked by λ repressor provided *in trans* (79). The resulting transcript includes nut_L, the site at which N protein can interact with RNA polymerase to enable transcription to proceed beyond stop signals. A coding sequence inserted at the HpaI site will be expressed if it includes a ribosome binding site. N protein can be provided *in trans* following derepression of the defective cro^- prophage that encodes repressor. In the presence of N protein transcription will read through the termination signal t_L. Middle: Transcription in pCVQ2 is driven from p_R, and repressor is provided by the cI gene (cI857) present on the plasmid (75). The cI gene is transcribed from p_{RM}, a promoter that is autogenously regulated by λ repressor. The λ sequence includes the Shine–Dalgarno sequence and initiation codon of the *cro* gene. Bottom: Transcription in pQTE is driven from the late λ promoter (22). *Q* gene product is essential for transcription from p_R' to proceed through t_R'. Derepression of the *lac* promoter activates transcription of *Q*. Coding sequences inserted downstream of p_R' will be translated if they have a ribosome binding site.

These strains are sensitive to UV light and grow poorly. Recent cro^- lysogens retain *uvrB* (69). They are more vigorous and respond better to the presence of N protein, though it is not possible to predict whether induction of the N^+ or N^- prophage will give the better amplification of protein (69).

A potential problem with the use of temperature induction is the concomitant induction of the heat-shock response, a response that includes the activation of proteolytic activity, which can therefore result in degradation of the foreign protein (3, 9). In addition, high temperature may have detrimental effects on the structure and solubility of protein products. Other methods of relieving repression have therefore been sought. For a wild-type *c*I gene this can be achieved by treatment with mitomycin C or nalidixic acid (85). Alternatively, λ repressor can be provided by a chromosomally located *c*I gene whose expression is dependent on a promoter taken from the *trp* operon of *Salmonella* (49, 56). In this case, the addition of tryptophan represses transcription of *c*I from p_{trp}, thereby leading to activation of transcription from p_L. Wild-type repressor provides very tight control of transcription from the powerful p_L promoter, tighter control than that provided by the thermosensitive form, while derepression is readily achieved in the absence of the heat-shock response.

Promoter p_R

The second plasmid shown in figure 43-2, pCVQ, incorporates p_R rather than p_L (75). The *c*I gene encoding a thermosensitive repressor is included within the plasmid vector and is transcribed from its normal promoter, p_{RM}, transcription from which is itself activated by CI protein (73). This autogenous regulation serves to maintain transcription of *c*I, but it should be noted that if this plasmid enters a cell that is devoid of CI, then transcription from p_{RM} will be inefficient and that from p_R will be efficient. This potential problem is readily avoided by the use of a host that is lysogenic for wild-type λ. The lysogenic strain can be cured of its prophage once the plasmid is established.

Promoter p_R'

Two plasmids have been described in which expression relies on the Q-dependent transcription from the late λ promoter p_R' (22, 71). This system is predicted to provide tight control from an efficient promoter using RNA polymerase modified by Q protein (99). As already mentioned, p_R' has been used effectively even within the context of a λ vector in which transcription must traverse approximately 20 genes before reading the coding sequence of interest (55, 63). In pQTE1, the third vector shown in figure 43-2, Q protein is supplied under the auspices of the control system of the *lac* operon (22). An alternative vector, pCEQ3, has transcription of gene *Q* under the control of a thermosensitive repressor (71). For this vector it was shown that induction of Q was achieved at 40°C in the absence of the heat-shock response. The potential promise of these systems was supported by the expression of *lamB* (22) and the IFNα2 gene (71), but recent literature appears to provide little evidence for the application of p_R'. It is probable that this system entered the scene when others were already refined and their use well established.

Translation of Messenger RNA

Transcription from p_L in plasmid vectors is seldom a problem. Translation, however, cannot be guaranteed; the secondary structure of the messenger RNA and consequent exposure of the ribosome binding site (RBS) are uncertain (40). Translation of messenger RNA from λ promoters can rely on a RBS associated with the cloned gene. Alternatively, the coding sequence can be positioned to take advantage of a RBS (e.g., gene *N*, *cro* or *c*II) within the vector. In some cases, a series of fusions is made in which the coding sequence is placed close to a Shine–Dalgarno sequence within the vector and the fusion that gives the highest yield of gene product is selected (78). Alternative systems with a higher probability of efficient translation are those designed to yield the foreign protein as a fusion product. A special example of this is the *E. coli* thioredoxin system in which the p_L promoter is controlled by the tryptophan-mediated repression of the *c*I gene and the thioredoxin part of the fusion protein directs the product to adhesion zones. From there it may be released by osmotic shock (49). Such fusion systems are likely to conserve a messenger RNA sequence in which the RBS is exposed for efficient translation.

Some Additional Modifications

Many expression plasmids use the early promoters of λ (5, 35, 48, 75, 77, 79). Some, such as pCYTEXP, include both in tandem (4). pCYTEXP is most notable for its inclusion of a sequence shown to favor efficient translation of the bacterial gene, *atpE*. This sequence was named TIR for Translation Initiation Region and is presumed to present the RBS within a favorable secondary structure. The pCYTEXP vector was made so that its components can be removed and exchanged. This modular construction means that the segment including TIR can be excised and its sequence subjected to site-directed mutagenesis with the aim of changing the sequence to enhance translation of the relevant foreign gene.

Another vector includes the *par* locus from plasmid pSC101 to ensure efficient segregation of the plasmid at cell division (51), a feature that may be advantageous in large-scale cultures.

E.coli Plasmids that Harbor a Promoter from a Virulent Phage

The Promoters of T3, T7, and SP6

Phage λ was an early paradigm for the study of gene regulation, and it is not surprising that its promoters, along with those of the *lac* and *trp* operons, were the first to be adopted for gene expression in plasmid vectors. The promoters of some virulent phages offer quite different approaches for controlled transcription. Most virulent phages possess powerful promoters and, in many phages, these include promoters that are not recognized by bacterial RNA polymerase. Phages T3 and T7 of *E. coli*, and their relative SP6 of *Salmonella typhimurium*, specify their own RNA polymerase each of which recognizes a phage-specific promoter sequence (see chapter 20 for review of phage T7 and its relatives). Only rarely do these phage polymerases encounter a "promoter-like" sequence in bacterial DNA. Also, these very selective RNA polymerases are more active than their bacterial counterparts and are seldom impeded by termination signals (89). The phage-specific polymerases were first used for the synthesis of RNA in vitro. Of the three phages mentioned, the RNA polymerase of SP6 was especially amenable because this enzyme is stable and easy to purify (11), but soon the gene specifying T7 RNA polymerase had itself been cloned and overexpressed. This enzyme also became easy to purify in large quantities (17, 91). An SP6 or T7 promoter within plasmid DNA is essentially the sole target for transcription by its cognate RNA polymerase (11, 52).

Such systems were quickly exploited for the synthesis in vitro of biologically active RNA (54). The resultant single-stranded RNA can serve, for example, as a transcript for translation, as antisense RNA, as a substrate for RNA processing or simply for the generation of labeled probes. Modifications to the majority of cloning vectors, whether plasmid, λ, or cosmid, were made so that their cloning sites were flanked by two promoters, either T7 and SP6 or T7 and T3, hence either strand of any cloned DNA can be transcribed in vitro.

The T7 promoter first used by Tabor and Richardson (91) and by Studier and Moffat (88) has probably become the most commonly used promoter for the amplification of gene products. Families of vectors, designated pET (Plasmid for Expression by T7 RNA polymerase) devised by

Studier, Rosenberg, and colleagues at Brookhaven National Laboratory (80, 89) have continued to evolve within the biotech company, Novagen.

The pET Vectors

The original pET vectors were derived from pBR322 and all included the promoter that precedes gene *10* of T7 (ϕ10) followed by unique sites at which a coding sequence can be introduced, so-called cloning sites (figure 43-3). Downstream of the cloning site(s) some vectors include Tϕ, the transcription terminator that follows gene *10* in the T7 genome. Others specify, instead, an RNase III cleavage site that does not cause termination of transcription but allows RNase III to process the transcript (80).

Some vectors lack a RBS and are essentially transcription vectors. Others are designed to favor efficient translation and to do so they take advantage of the translation start signals of gene *10*, referred to as S10 (see figure 43-3). The product of gene *10*, the major capsid protein of T7, is made more rapidly than any other phage protein during infection, and its RBS is believed to be set within the context of an RNA sequence that favors translation (58, 66). In the early vectors, in-frame fusions could be made at early or late positions within gene *10* (codon 11 or 260); RNA from such fusions is likely to retain the secondary structure that favors translation of the gene *10* sequence. Alternative vectors include an NdeI site within the second and third codons and are used to join coding sequences without making changes within the sequence of the foreign protein. However, the efficiency of translation of the mRNA from these fusions is less predictable because the new sequence may change the secondary structure of the RNA.

More recent vectors are modified to generate gene products fused to any of a variety of tags. The tag may be chosen to aid identification, purification or stability of the product. They may also include a sequence that encodes a target for proteolysis to facilitate separation of the protein of interest from its tag.

Successful control of expression from the T7 promoter within the pET vectors is determined by the availability of the T7 RNA polymerase. Classically, the provision of the T7 RNA polymerase specified by a λ prophage (λDE3) has been under the control of the *lac*UV5 promoter. Induction of transcription from the T7 promoter follows the addition of IPTG and the consequent expression of the T7 RNA polymerase gene within the resident prophage. This system of control provides effective transcription from the T7 promoter in the pET vector but allows some constitutive (basal) transcription of the RNA polymerase

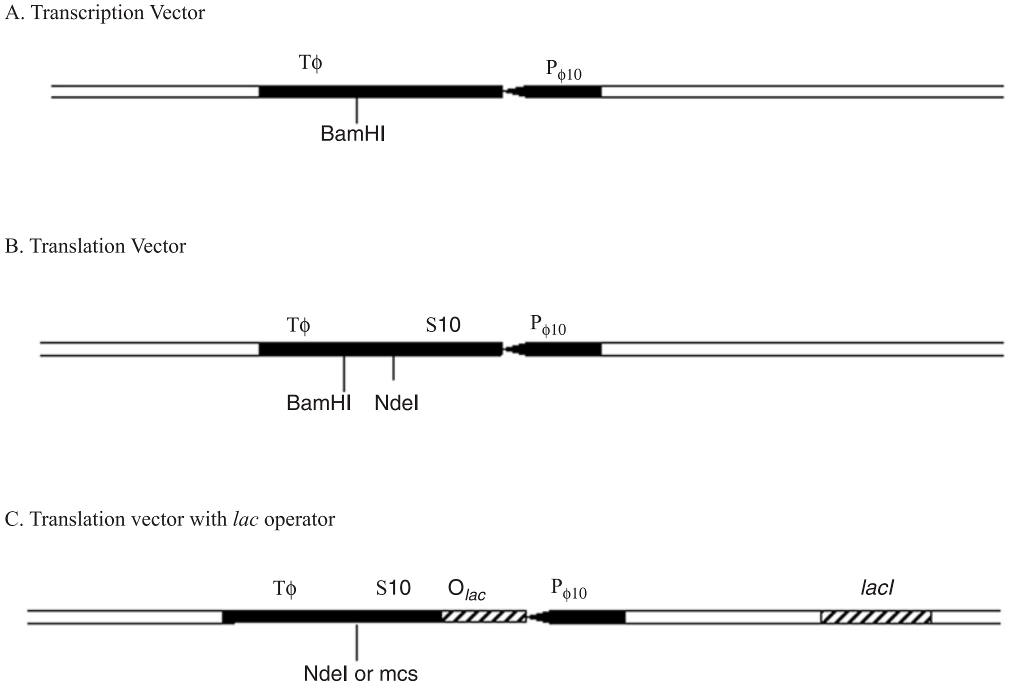


Figure 43-3 The key features of pET vectors. In each diagram the relavent elements of T7 are located within the black segment. All pET vectors are based on the promoter of gene *10* of phage T7 ($P_{\phi 10}$) (88). Some include the ribosome binding site (S10) of this gene and those shown include the transcription terminator, Tϕ, downstream of gene *10*. In the simple translation vector the NdeI site permits fusion of a coding sequence to the ribosome binding site of gene *10*, while insertion at the BamHI site generates a fusion protein. More recent vectors include a series of targets for alternative restriction enzymes — a multiple cloning site (MCS).

gene. Even a low level of RNA polymerase can be a grave problem if the foreign gene encodes a protein that is toxic to the bacterium. Various modifications have been used to tighten the control of expression from the T7 promoter. One modification depends on the presence of T7 lysozyme, which has been shown to be a natural inhibitor of T7 RNA polymerase (39, 58) (see chapter 20). In some cases, T7 lysozyme, specified on a compatible plasmid, can suffice to reduce the basal activity of T7 RNA polymerase in the lysogenic strain and allow potentially toxic genes to be established in the context of the pET vector (87). It is possible, though less convenient, to maintain the foreign gene in the complete absence of T7 RNA polymerase and activate expression of the foreign gene by infection with a λ phage (λ CE6) that encodes the T7 RNA polymerase.

An additional refinement of pET vectors is the inclusion of a *lac* operator in association with the T7 promoter, hence *p*T7*lac* (19). The basal level of transcription from the T7 promoter is reduced by a *lac* operator centered 15 bp downstream from the RNA start. In the absence of the bound repressor, or following the addition of inducers, transcription from the T7 promoter remains powerful. The new pET vectors carry a *lacI* gene to provide sufficient repressor to control both the chromosomal gene for T7 RNA polymerase and the T7*lac* promoter in the multicopy expression vector. The pET vectors based on the T7*lac* promoter and the *lacI* gene (figure 43-3) are able to maintain most genes in the presence of the λ DE3 prophage, even if they encode a toxic gene product. In this system, the basal level of expression from the T7 promoter is very low and yet induction can produce high levels of the toxic gene product. In some cases, where this low level of residual expression makes the plasmid difficult to maintain, the addition of a compatible plasmid encoding T7 lysozyme has stabilized the system. More recently, the dual control of plasmid copy number and transcription serves to block the production of toxic gene products prior to induction (97). The maintenance of pETcoco (a copy control derivative of pT7*lac*) in single copy, rather than 30 copies, provides a proportional (approximately 30-fold) decrease in the basal level of transcription from the T7 promoter.

T7 Promoters in Organisms other than E. coli

The T7 system has been used effectively in streptomycetes (36, 72). In this case an inducible T7 RNA polymerase gene has been placed within the bacterial chromosome (J. Altenbuchner, personal communication). Expression of foreign genes from a T7 promoter has been demonstrated in the cytoplasm of mammalian cells following delivery by a viral vector (31) and there is evidence that T7 RNA polymerase can be directed to the nucleus by the inclusion of a nuclear location signal (20). When T7 RNA polymerase is overproduced in the cytoplasm of yeast, some enters the nucleus and elicits transcription from a T7 promoter sequence in the chromosome (14).

Other Coliphage Promoters

Some virulent phages, including T5 and the T-even phages (see chapters 19 and 18, respectively), do not specify their own RNA polymerase. Instead, these phages impose modifications that subvert the transcription machinery of their hosts. The modifications are not easily exploited, but both the extraordinary strength of a T5 promoter and features that enhance translation of messenger RNA from a T4 promoter have been tested in expression vectors. For instance, a synthetic T5 promoter in an association with an effective RBS has served to produce human interferon-γ in *E. coli*. In this example, successful amplification of the gene product was obtained without any control of transcription, and interferon constituted 15% of total cell protein (42).

The *p*N25 promoter of T5 outcompetes the promoter for the β-lactamase gene by a ratio of 25:1 (90). A very versatile plasmid vector (pDS6) includes this promoter fused to the *lac* operator (*p*N25x/o) in the absence of a RBS. The cloning sites separate *p*N25x/o from a transcription termination signal taken from phage λ. In vivo, transcription from this promoter can be moderated by *lac* repressor, and translation of messenger RNA can be obtained if the cloned coding sequence specifies a RBS. The power of the T5 promoter dictates that approximately 95% of the transcripts from plasmid DNA initiate from the T5 promoter. In vitro, in the presence of 7mGpppA and the four NTPs, the resulting messenger RNA is capped and serves directly for translation in a wheat germ system (90).

One plasmid expression system derives from phage T4. During the life cycle of this phage, massive quantities of gene *32* product are required. Gene *32* is endowed with a strong promoter and its messenger RNA includes efficient signals for translation. The messenger RNA is noteworthy not only because it is translated with high efficiency but also because it is unusually stable. Using these features in an expression vector (21) it was shown that infection with phage T4 could enhance the stability of a foreign protein. At least part of this protective effect is dependent on the *pin* gene of T4. The Pin protein inhibits the Lon protease and may inhibit other proteases (86). Pin can be provided *in trans* from a compatible plasmid.

Convergent Transcription

RNA polymerase complexes traveling from opposing promoters may collide if no transcription termination signal intervenes. Convergent transcription can impair gene expression, more particularly that from the weaker promoter. This inhibition of gene expression can result from the interaction of messenger RNA with antisense RNA (16) or from the *cis*-specific, steric hindrance of transcription (23, 95). Within a natural context, convergent transcription can make a regulatory contribution to a developmental pathway, as documented for phage λ (see, for example, 46, 53). Within the context of an expression vector, uncontrolled transcription from a strong promoter could interfere with the expression of an essential gene within the plasmid and possibly with the replication of the vector (7, 77). On the other hand, transcriptional interference has been used constructively to tighten the repression of a gene specifying a restriction endonuclease in a cell lacking protection from the cognate modification enzyme (65). This principle (figure 43-4) was proven to enable the cloning of a gene encoding a restriction endonuclease in the absence of the modification gene (44). The converse effect — the relief of transcriptional interference and the consequent expression of the gene from the opposing, weaker, promoter — was developed as a selection system for sequence-specific DNA-binding proteins (24). In this situation expression from the weaker promoter was recognized by the acquisition of a drug-resistant phenotype when a sequence-specific DNA-binding protein blocks transcription from the opposing promoter.

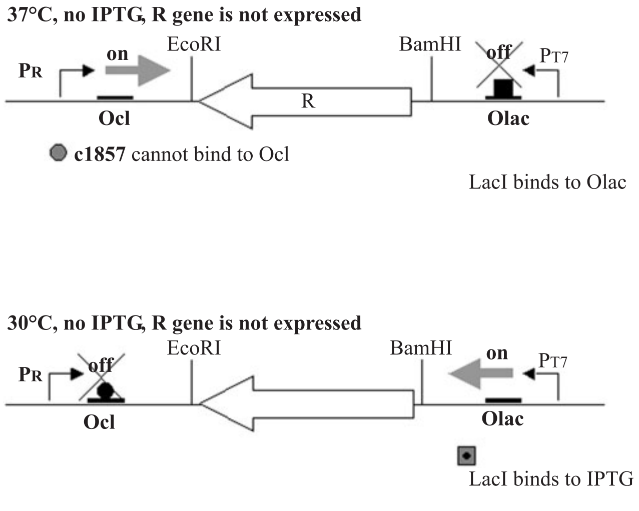


Figure 43-4 Plasmid pLT7K and its use. An illustration of the use of pLT7K to clone a gene (*R*) specifying a restriction endonuclease in the absence of its cognate modification enzyme. At 37°C in the absence of IPTG the Lac repressor binds to the *lac* operator, and transcription from the T7 promoter is minimal; the λ temperature-sensitive repressor is inactivated, and transcription from "P_R" (the λ promoter, p_R) antagonizes any residual transcription of gene *R* from the T7 promoter. At 30°C, in the presence of IPTG, transcription of gene *R* occurs from the T7 promoter, and transcription from the opposing promoter is repressed. This figure is taken with permission from Kong et al. (44).

Noncoliphage Promoters

Bacillus subtilis is an extensively studied Gram-positive bacterium and an attractive host for protein production for a number of reasons: it is nonpathogenic, its fermentation properties are well understood, and it is capable of secreting proteins, especially those originating from Gram-positive bacteria. A number of *B. subtilis* phages have been well studied and some advantage has been taken of their promoters. The late promoter (A3) of bacteriophage φ29, maintained within a plasmid vector, provides constitutive expression during exponential growth of *Streptomyces lividans* (74). This system failed to give high levels of agarase in *Streptomyces* species (68). A hybrid control system, derived from a promoter of the virulent *B. subtilis* phage SP01 and the *lac* operator of *E. coli*, has been used in *B. subtilis* to give inducible expression of a human leukocyte interferon A gene (98).

Considerable effort has been given to harnessing an inducible promoter of a *B. subtilis* prophage in order to take advantage of maintenance as a single copy prior to induction (32, 50, 93). The "shot-gun" cloning of a promoter-less *lacZ* gene in the temperate phage φ105 identified a strong inducible promoter. Transcription of heterologous genes, inserted at the site identified by the reporter, can be controlled by a temperature-sensitive repressor and, fortuitously, the insertion itself makes the prophage defective in cell lysis. The site of insertion appears to be functionally equivalent to that downstream of p_R' in phage λ.

A strong transcriptional element comprising a pair of tandemly arranged promoters from phage I19 of *Streptomyces ghanaens* has been isolated and shown to respond to RNA polymerase holoenzyme from *E. coli* and *S. lividans* (47). However, as mentioned earlier, the T7 promoter system is now available in streptomycetes (72).

E. coli still remains the chief bacterial factory for the production of proteins within the laboratory. Nevertheless, phage promoters continue to be exploited in a wide variety of bacteria for many genetic analyses (94) and tests, including their use with the luciferase reporter within mycobacteriophages as a sensitive assay for live mycobacteria (41, 82) (these methods are additionally reviewed in chapters 38 and 46, covering mycobacteriophage and phage-based bacterial diagnosis, respectively).

A Phage-Based Perspective

The amplification of heterologous proteins is still an unpredictable mission: efficient translation of messenger RNA is not guaranteed, the protein may fail to assemble correctly, and the protein may be a substrate for degradation by host proteases. Expression of coding sequences as fusion products may overcome these problems.

Control of transcription has always been of prime importance for protein amplification. Plasmid systems dependent on an early promoter of λ or the T7 gene *10* promoter have been refined to combat this problem, but phage λ and its prophage remain a simple though under-used alternative. Early experiments with *polA* illustrate the vulnerability of *E. coli* to the expression of genes in a multicopy plasmid. The *polA* gene with its own promoter is stably maintained in a λ vector during lytic or lysogenic propagation, but could not be maintained in a multicopy plasmid (43, 64). When the promoter-defective gene was transferred to a multicopy plasmid, with transcription of *polA* from the p_L promoter of λ controlled by temperature-sensitive repressor, the recombinant plasmid remained difficult to maintain (57).

The role of Cro in the moderation of the early promoter of phage λ precludes the optimal use of p_L from an induced prophage. However, the late promoter p_R' has been harnessed effectively even when shared with most of the late genes of λ; T4 DNA ligase and Pnk comprised 5% and 7% of soluble cell protein, respectively (55, 63). The expression of a heterologous gene from p_R' of a prophage requires derepression of p_R to produce Q protein, which is necessary for expression of genes from p_R'. Could the prophage system be modified to take even better advantage of p_R' than currently reported? It probably could. A genetic trick to make the prophage defective by deletion of the 20 or more phage genes that separate the heterologous gene from p_R' would focus the protein-synthesizing machinery on the message of interest, and the amplification of the heterologous product would no longer be accompanied by the production of many phage proteins. Both proximity to the p_R' promoter and loss of competition from 20 phage genes should elevate amplification of the protein of interest. However, the new "copy control" plasmids now offer the same benefits as prophage induction. Such plasmids provide tight control from a cloned gene in single copy, and efficient transcription following plasmid amplification (84, 97).

Acknowledgments

I am indebted to Alexander Gann, Kenneth Murray, and Waclaw Szybalski for their constructive comments on the manuscript, to Jeffrey Errington, Rafa Mellado, Graham Hatfull, Thomas Trautner, and Wolfgang Wohlleben for helping me access the literature for noncoliphages, to Joe Altenbuchner for permission to cite his unpublished work, to Mary O'Neill for artwork, and to Alix Fraser for preparing the manuscript.

References

1. Adhya, S., M. Gottesman, and B. de Crombrugghe. 1974. Release of polarity in *Escherichia coli* by gene *N* of phage λ: termination and antitermination of transcription. Proc. Natl. Acad. Sci. USA 71:2534–2538.
2. Allgood, N. D., and T. J. Silhavy. 1988. Illegitimate recombination in bacteria, pp. 309–330. *In* R. Kucherlapati and G. R. Smith (eds.) Genetic Recombination. American Society for Microbiology, Washington, D.C.
3. Baker, T. A., A. D. Grossman, and C. A. Gross. 1984. A gene regulating the heat-shock response in *Escherichia coli* affects proteolysis. Proc. Natl. Acad. Sci. USA 81:6779–6783.
4. Belev, T. N., M. Singh, and J. E. G. McCarthy. 1991. A fully modular vector system for the optimization of gene expression in *Escherichia coli*. Plasmid 26:147–150.
5. Bernard, H.-U., and D. R. Helsinski. 1979. Use of the lambda phage promoter p_L to promote gene expression in hybrid plasmid cloning vehicles. Methods Enzymol. 68:482–492.
6. Brammar, W. J. 1977. The construction in vitro and exploitation of transducing derivatives of bacteriophage λ. Biochem. Soc. Trans. 5:1633–1652.
7. Brewer, B. J. 1988. When polymerases collide: replication and the transcriptional organization of *E. coli*. Cell 53:679–686.
8. Bøvre, K., and W. Szybalski. 1969. Pattern of convergent and overlapping transcription within the *b*2 region of coliphage λ. Virology 38:614–626.
9. Buell, G., M.-E. Schultz, G. Selzer, A. Chollet, N. R. Movva, D. Semond, S. Escanez, and E. Kawashima. 1985. Optimizing the expression in *E. coli* of a synthetic gene encoding somatomedin-C (IGF-I). Nucleic Acids Res. 13:1923–1938.
10. Burt, D. W., and W. J. Brammar. 1982. Transcriptional termination sites in the *b*2 region of bacteriophage lambda that are unresponsive to antitermination. Mol. Gen. Genet. 185:462–467.
11. Butler, E. T., and M. J. Chamberlin. 1982. Bacteriophage SP6-specific RNA polymerase. I. Isolation and characterization of the enzyme. J. Biol. Chem. 257:5772–5778.
12. Campbell, A. 1962. Episomes. Adv. Genet. 11:101–145.
13. Castellazzi, M., P. Brachet, and H. Eisen. 1972. Isolation and chracterization of deletions in bacteriophage lambda residing as prophage in *E. coli*. Mol. Gen. Genet. 117:211–218.
14. Chen, W., S. Tabor, and K. Struhl. 1987. Distinguishing between mechanisms of eukaryotic transcriptional activation with bacteriophage T7 RNA polymerase. Cell 50:1047–1055.
15. Cohen, S. N., A. C. Y. Chang, H. W. Boyer, and R. Helling. 1973. Construction of biologically functional plasmids in vitro. Proc. Natl. Acad. Sci. USA 70:3240–3244.
16. Coleman, J., P. J. Green, and M. Inouye. 1984. The use of RNAs complementary to specific mRNAs to regulate the expression of individual bacterial genes. Cell 37:429–436.
17. Davanloo, P., A. H. Rosenberg, J. J. Dunn, and F. W. Studier. 1984. Cloning and expression of the gene for bacteriophage T7 RNA polymerase. Proc. Natl. Acad. USA 81:2035–2039.
18. Davison, J., W. J. Brammar, and F. Brunel. 1974. Quantitative aspects of gene expression in a λ-*trp* fusion operon, Mol. Gen. Genet. 130:9–20.
19. Dubendorff, J. W., and F. W. Studier. 1991. Controlling basal expression in an inducible T7 expression system by

blocking the target T7 promoter with *lac* repressor. J. Mol. Biol. 219:45–49.

20. Dunn, J. J., B. Krippl, K. E. Bernstein, H. Westphal, and F. W. Studier. 1988. Targeting bacteriophage T7 RNA polymerase to the mammalian nucleus. Gene 68:259–266.
21. Duvoisin, R., D. Belin, and H. M. Krisch. 1986. A plasmid expression vector that permits stabilization of both mRNAs and proteins encoded by cloned genes. Gene 45:193–201.
22. Edlind, T., R. Young, R. Eller, T. Hernandez, and G. Ihler. 1986. Plasmid expression vector using the lambda late promoter. Plasmid 15:242–244.
23. Elledge, S. J., and R. W. Davis. 1989. Position and density effects on repression by stationary and mobile DNA-binding proteins. Genes Dev. 3:185–197.
24. Elledge, S. J., P. Sugiono, L. Guarente, and R. W. Davis. 1989. Genetic selection for genes encoding sequence-specific DNA-binding proteins. Proc. Natl. Acad. Sci. USA 86:3689–3693.
25. Folkmanis, A., Y. Takeda, J. Simuth, G. Gussin, and H. Echols. 1976. Purification and properties of a DNA-binding protein with characteristics expected for the Cro protein of bacteriophage λ, a repressor essential for lytic growth. Proc. Natl. Acad. Sci. USA 73:2249–2253.
26. Folkmanis, A., W. Maltzman, R. Mellon, A. Skalka, and H. Echols. 1977. The essential role of the *cro* gene in lytic development by bacteriophage λ. Virology 81:352–362.
27. Franklin, N. C. 1971. Illegitimate recombination, pp. 175–194. *In* A. D. Hershey (ed.) The Bacteriophage Lambda. Cold Spring Harbor Laboratories, Cold Spring Harbor, N.Y.
28. Franklin, N. C. 1971. The *N* operon of lambda: extent and regulation as observed in fusions to the tryptophan operon of *Escherichia coli*, pp. 621–638. *In* A. D. Hershey (ed.) The Bacteriophage Lambda. Cold Spring Harbor Laboratories, Cold Spring Harbor, NY.
29. Franklin, N. C. 1974. Altered reading of genetic signals fused to the *N* operon of bacteriophage λ: genetic evidence for modification of polymerase by the protein product of the *N* gene. J. Mol. Biol. 89:33–48.
30. Franklin, N. C. 1978. Genetic fusions for operon analysis. Annu. Rev. Genet. 12:193–221.
31. Fuerst, T. R., E. G. Niles, F. Studier, and B. Moss. 1986. Eukaryotic transient-expression system based on recombinant vaccinia virus that synthesizes T7 RNA polymerase. Proc. Natl. Acad. Sci. USA 83:8122–8126.
32. Gibson, R. M., and J. Errington. 1992. A novel *Bacillus subtilis* expression vector based on bacteriophage phi105. Gene 121:137–142.
33. Gottesman, M., S. Adhya, and A. Das. 1980. Transcription antitermination by bacteriophage λ *N*-gene product. J. Mol. Biol. 140:57–75.
34. Harris, A. W., D. W. A. Mount, C. R. Fuerst, and L. Siminovitch. 1967. Mutations in bacteriophage λ affecting host cell lysis. Virology 32:553–569.
35. Hedgpeth, J., M. Ballivet, and H. Eisen. 1978. Lambda phage promoter used to enhance expression of a plasmid-cloned gene. Mol. Gen. Genet. 163:197–203.
36. Heinzelmann, E., G. Kienzlen, S. Kaspar, J. Recktenwald, W. Wohlleben, and D. Schwartz. 2001. The phosphoinomethylmalate isomerase gene *pmi*, encoding an aconitase-like enzyme, is involved in the synthesis of phosphinothricin tripeptide in *Streptomyces viridochromogenes*. Appl. Environ. Microbiol. 67:3603–3609.
37. Honigman, A. 1981. Cloning and characterisation of a transcription termination signal in bacteriophage λ unresponsive to the *N* gene product. Gene 13:299–309.
38. Hopkins, A. S., N. E. Murray, and W. J. Brammar. 1976. Characterisation of λ*trp*-transducing bacteriophages made in vitro. J. Mol. Biol. 107:549–569.
39. Inouye, M., N. Arnheim, and R. Sternglanz. 1973. Bacteriophage T7 lysozyme is an N-acetylmuramyl-L-alanine amidase. J. Biol. Chem. 248:7247–7252.
40. Iserentant, D., and W. Fiers. 1980. Secondary structure of mRNA and efficiency of translation initiation. Gene 9:1–12.
41. Jacobs, W. R., R. Udani, R. Barletta, J. Chan, G. Kalkut, G. Sonse, T. Kieser, G. Sarkis, G. F. Hatfull, and B. R. Bloom. 1993. Rapid assessment of drug susceptibilities of *Mycobacterium tuberculosis* by means of luciferase reporter phages. Science 260:819–822.
42. Jay, E., J. Tommens, L. Pomeroy-Cloney, D. MacKnight, C. Lutz-Wallace, P. Wishart, D. Harrison, W.-Y. Lui, V. Asundi, M. Dawood, and F. Joy. 1984. High-level expression of a chemically synthesized gene for human interferon-γ using a prokaryotic expression vector. Proc. Natl. Acad. Sci. USA 81:2290–2294.
43. Kelley, W. S., K. Chalmers, and N. E. Murray. 1977. Isolation and characterisation of a λ*polA* transducing phage. Proc. Natl. Acad. Sci. USA 74:5632–5636.
44. Kong, H., L.-F. Lin, N. Porter, S. Stickel, D. Byrd, J. Posfai, and R. J. Roberts. 2000. Functional analysis of putative restriction-modification system genes in the *Helicobacter pylori* J99 genome. Nucleic. Acids Res. 28:3216–3223.
45. Krell, K., M. E. Gottesman, J. S. Parks, and M. A. Eisenberg. 1972. Escape synthesis of the biotin operon in induced λ*b2* lysogens. J. Mol. Biol. 68:69–82.
46. Krinke, L., M. Mahoney, and D. L. Wulff. 1991. The role of the OOP antisense RNA in coliphage development. Mol. Microbiol. 5:1265–1272.
47. Labes, G., M. Bibb, and W. Wohlleben. 1997. Isolation and characterization of a strong promoter element from *Streptomyces ghanaensis* phage I19 using the gentamicin resistance gene (aacC1) of Tn*1696* as reporter. Microbiology. 143:1503–1512.
48. Lautenberger, J. A., D. Court, and T. S. Papas. 1983. High-level expression in *Escherichia coli* of the carboxyl-terminal sequences of the avian myelocytomatosis virus (MC29) *v-myc* protein. Gene 23:75–84.
49. LaVallie, E. R., E. A. Di Blasio, S. Kovacic, K. L. Grant, P. E. Schendel, and J. J. McCoy. 1993. A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm. Bio/Technology 11:187–193.
50. Leung, Y.-C., and J. Errington. 1994. Characterization of an insertion in the phage phi105 genome that blocks host

cell lysis and provides strong expression of heterologous genes. Gene 154:1–6.

51. Love, C.A., P. E. Lilley, and N. E. Dixon. 1996. Stable high-copy number bacteriophage λ promoter vectors for the overproduction of proteins in *E. coli*. Gene 176:49–53.
52. McAllister, W. T., C. Morris, A. H. Rosenberg, and W. E. Studier. 1981. Utilisation of bacteriophage T7 late promoters in recombinant plasmids during infection. J. Mol. Biol. 153:527–544.
53. McMacken, R., N. Mantei, B. Butler, A. Joyner, and H. Echols. 1970. Effect of mutations in the *c*II and *c*III genes of bacteriophage λ on macromolecular synthesis in infected cells. J. Mol. Biol. 49:639–655.
54. Melton, D. A., P. D. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green. 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic. Acids Res. 12:7035–7056.
55. Midgley, C. A., and N. E. Murray. 1985. T4 polynucleotide kinase: cloning of the gene (*pse*T) and amplification of its product. EMBO J. 19:2695–2703.
56. Mieschendahl, M., T. Petri, and U. Hänggi. 1986. A novel prophage independent *trp*-regulated lambda p_{L} expression system. Bio/technology 4:802–808.
57. Minkley, E. G., A. T. Leney, J. B. Bodner, M. P. Panicker, and W. E. Brown. 1984. *Escherichia coli* DNA polymerase I. Construction of a *polA* plasmid for amplification and an improved purification scheme. J. Biol. Chem. 259:10386–10392.
58. Moffatt, B. A., and F.W. Studier. 1987. T7 lysozyme inhibits transcription by T7 RNA polymerase. Cell 49:221–227.
59. Moir, A., and W. J. Brammar. 1976. The use of specialised transducing phages in the amplification of enzyme production. Mol. Gen. Genet. 149:87–99.
60. Müller-Hill, B., L. Crapo, and W. Gilbert. 1968. Mutants that make more *lac* repressor. Proc. Natl. Acad. Sci. USA 59:1259–1264.
61. Müller-Hill, B. 1975. *Lac* repressor and *lac* operator. Prog. Biophys. Mol. Biol. 30:227–252.
62. Murray, N. E., and K. Murray. 1974. Manipulation of restriction targets in phage λ to form receptor chromosomes for DNA fragments. Nature 252:476–481.
63. Murray, N. E., S. A. Bruce, and K. Murray. 1979. Molecular cloning of the DNA ligase gene from bacteriophage T4. II. Amplification and purification of the gene product. J. Mol. Biol. 132:493–505.
64. Murray, N. E., and W. S. Kelley. 1979. Characterisation of λ*polA* transducing phages: effective expression of the *E. coli polA* gene. Mol. Gen. Genet. 175:77–87.
65. O'Connor, C. D., and K. N. Timmis. 1987. Highly repressible expression system for cloning genes that specify potentially toxic proteins. J. Bacteriol. 169:4457–4462.
66. Olins, P. O., C. S. Devine, S. H. Rangwala, and K. S. Kavka. 1988. The T7 phage gene *10* leader RNA, a ribosome-binding site that dramatically enhances the expression of foreign genes in *E. coli*. Gene 73:227–235.
67. Panasenko, S. N., J. R. Cameron, R. W. Davis, and I. R. Lehman. 1977. Five-hundred fold overproduction of DNA ligase after induction of a hybrid lambda lysogen constructed in vitro. Science 196:188–189.
68. Parro, V., C. Vives, F. Godia, and R. P. Mellado. 1997. Overproduction of an agarase of bacterial origin. J. Biotech. 55:59–66.
69. Patterson,T. A., N. Costantino, S. Dasgupta, and D. L. Court. 1993. Improved hosts for regulated expression of genes from p_{L} plasmid vectors. *Gene* 132:83–87.
70. Pero, J. 1971. Deletion mapping of the site of action of the *tof* gene product, pp. 599–608. *In* A. D. Hershey (ed.) The Bacteriophage Lambda. Cold Spring Harbor Laboratories, Cold Spring Harbor, N.Y.
71. Petrenko, L. A., I. P. Gileva, and V. V. Kravchenko. 1989. Expression vector with two-step control by the *c*I-p_{R}-*Q*-$\mathrm{p_R}'$*qut*-t'_{R} module of coliphage lambda. Gene 78:85–91.
72. Pfeifer, V., G. J. Nicholson, J. Ries, J. Recktenwald, A. B. Schefer, R. Shawky, J. Schröder, W. Wohlleben, and S. Pelzer. 2001. A polyketide synthase in glycopeptide biosynthesis: the biosynthesis of the non-proteinogenic amino acid (*S*)-3,5-dihydroxyphenylglycine. J. Biol. Chem. 276:38370–38377.
73. Ptashne, M. 1992. *In* The Genetic Switch. Cell Press and Blackwell Scientific, Oxford.
74. Pulido, D., A. Jiménez, M. S. Salas, and R. P. Mellado. 1986. *Bacillus subtilis* phage ϕ29 main promoters are effectively recognized in vivo by the *Streptomyces lividans* RNAp. Gene 49: 377–382.
75. Queen, C. 1983. A vector that uses phage signals for efficient synthesis of proteins in *Escherichia coli*. J. Mol. Appl. Genet. 2:1–10.
76. Rambach, A., and P. Tiollais. 1974. Bacteriophage λ having *Eco*RI endonuclease sites only in the non-essential region of the genome. Proc. Natl. Acad. Sci. USA 71:3927–3930.
77. Remaut, E., P. Stanssens, and W. Fiers. 1981. Plasmid vectors for high-efficiency expression controlled by the p_{L} promoter of coliphage lambda. Gene 15:81–93.
78. Roberts,T. M., R. Kacich, and M. Ptashne. 1979. A general method for maximizing the expression of a cloned gene. Proc. Natl. Acad. Sci. USA 76:760–764.
79. Rosenberg, M., Y.-S. Ho, and A. Shatzman. 1983. The use of pKC30 and its derivatives for controlled expression of genes, Methods Enzymol. 101:123–138.
80. Rosenberg, A. H., B. N. Lade, D.-S. Chui, S.-W. Lin, J. J. Dunn, and F. W. Studier. 1987. Vectors for selective expression of cloned DNAs by T7 RNA polymerase. Gene 56:125–135.
81. Salstrom, J. S., and W. Szybalski. 1978. Coliphage λ*nut*$_{\mathrm{L}}^{-}$: a unique class of mutants defective in the site of gene *N* product utilization for antitermination of leftward transcription. J. Mol. Biol. 124:195–221.
82. Sarkis, G., W. J. Jacobs, Jr., and G. F. Hatfull. 1995. L5 luciferase reporter mycobacteriophages: a sensitive tool for the detection and assay of live mycobacteria. Mol. Microbiol. 15:1055–1067.
83. Segawa, T., and F. Imamoto. 1974. Diversity of regulation of genetic transcription. II. Specific relaxation of polarity in read-through transcription of the translocated *trp* operon in bacteriophage λ*trp*. J. Mol. Biol. 87:741–754.

84. Sektas, M., and W. Szybalski. 2002. Novel single-copy pETcoco^{TM} vector with dual controls for amplification and expression. inNovations 14:6.
85. Shatzman, A. R., and Rosenberg, M. 1987. Expression, identification and characterization of recombinant gene products in *Escherichia coli*. Methods Enzymol. 152: 661–673.
86. Simon, L. 1994. T4-induced inhibition of proteolysis, pp. 382–384. *In* J. Karam (ed.) *Escherichia coli* in Molecular Biology of Bacteriophage T4. American Society for Microbiology, Washington, D.C.
87. Studier, F. W. 1991. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. J. Mol. Biol. 219:37–44.
88. Studier, F. W., and B. A. Moffatt. 1986. Use of bacteriophage T7 polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. 189:113–130.
89. Studier, F. W., A. H. Rosenberg, J. J. Dunn, and J. W. Dubendorff. 1990. Use of T7 RNA polymerase to direct expression of cloned genes. Method Enzymol. 185:60–89.
90. Stueber, D., L. Ibrahimi, D. Cutler, B. Dobberstein, and H. Bujard. 1984. A novel in vitro transcription-translation system: accurate and efficient synthesis of single proteins from cloned DNA sequences. EMBO J. 3:3143–3148.
91. Tabor, S., and C. C. Richardson. 1985. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. Proc. Natl. Acad. Sci. USA 82:1074–1078.
92. Thomas, M., J. R. Cameron, and R. W. Davis. 1974. Viable molecular hybrids of bacteriophage lambda and eukaryotic DNA. Proc. Natl. Acad. Sci. USA 71:4579–4583.
93. Thornewell, S. J., A. K. East, and J. Errington. 1993. An efficient expression and secretion system based on *Bacillus subtilis* phage ϕ105 and its use for the production of *B. cereus* β-lactamase I. Gene 133:47–53.
94. Vagner, V., E. Dervyn, and S. D. Ehrlich. 1998. A vector for systematic gene inactivation in *Bacillus subtilis*. Microbiology. 144:3097–3104.
95. Ward, D. F., and N. E. Murray. 1979. Convergent transcription in bacteriophage λ: interference with gene expression. J. Mol. Biol. 133:249–266.
96. Weisberg, R. A., and S. Adhya. 1977. Illegitimate recombination in bacteria and bacteriophage. Annu. Rev. Genet. 11:451–473.
97. Wild, J., and W. Szybalski. 2004. Copy-control tightly regulated expression vectors based on pBA/oriV, pp. 155–167. *In* P. Balbas and A. Lorence (eds.) Recombinant Gene Expression Reviews and Protocols. Methods in Molecular Biology, vol. 267 (J. M. Walker, Series Editor). Humana Press, Totowa, N.J.
98. Yansura, D. G., and D. J. Henner. 1984. Use of the *Escherichia coli lac* repressor and operator to control gene expression in *Bacillus subtilis*. Proc. Natl. Acad. Sci. USA 81:439–443.
99. Yarnell, W. S., and J. W. Roberts. 1992. The phage λ gene *Q* transcription antiterminator binds DNA in the late gene promoter as it modifies RNA polymerase. Cell 69: 1181–1189.

44

Phage in Display

BJORN H. LINDQVIST

Phage display is a process by which a peptide or a protein is expressed as an exterior fusion to a surface protein of a phage particle. The peptide or protein sequence can be deduced from its encoding DNA sequence that resides in the phage particle or in a transductant. Amplification of the DNA of interest can take place by phage/transductant propagation or by polymerase chain reaction (PCR). By producing large populations of phage particles, each expressing a unique peptide or protein, peptide/protein libraries can be obtained. Peptides or proteins, interacting with defined molecular targets — most often proteins — can be isolated from such libraries by enrichment through repeated cycles of panning. Hence, phage display can be thought of as a "search engine" of protein–target interactions.

The pioneering work of Smith (113) first demonstrated surface display of peptides in filamentous phage fd. This innovation was extended to peptide libraries of fd and M13 (19, 22, 108) and phagemid display was introduced (6). The display of proteins such as antibody domains and combinatorial antibody libraries soon followed (5, 59, 75). From the beginning of the 1990s phage-display-related publications have grown exponentially (128). Several reviews (some general, e.g., 58, 125; and many specialized) are available. There are numerous reports and several laboratory manuals (4) describing development and use of filamentous phage display in identification of peptide or protein interactions with simple organic compounds, antibodies, receptors, etc. Phage display is also a useful tool in protein engineering and directed evolution (44). Then there is the large sector of phage antibody display (64) and the more recent field of immune profiling and its implication for vaccine development (15, 112). Furthermore, complex targets such as cells (92) and whole tissues/organs (91) have been subjected to phage display analysis. These studies explore novel approaches for in vivo homing in gene/drug delivery (78), cancer surveillance/treatment (2, 86, 103), and imaging (127).

To extend the powers of filamentous phage display to other phage systems, phages λ, T7, and T4 have now been engineered for peptide/protein display (10, 49). This chapter will focus on developments in filamentous phage display and the emerging λ, T4, and T7 display systems. Current phage display options are listed in table 44-1.

Filamentous Phage Display

The biology of the filamentous single-stranded DNA phages (f1, fd, and M13) is described in chapter 12. For the sake of phage display, these particles consist of one major- and four minor-type coat proteins embedding the circular single-stranded DNA genome of 6400 nucleotides. The major coat protein (P8*) of phage M13 consists of 50 amino acids and is present in 2700 copies in the phage particle. Proteins P3 and P6, which are localized at the infecting tip of the particle, are present in five copies each while proteins P7 and P9 reside at the opposite end of the phage particles.

The particles infect *E. coli* F^+ cells by interacting with the F-pili through their minor capsid component, P3, leading to single-stranded DNA injection and conversion to a replication-form DNA (RF-DNA) molecule capable of rolling-circle DNA replication. In the case of phage fd, the P3 protein recognizes the primary receptor of infection, the F-pilus, through its N-terminal domain, D2, while the neighboring region, D1, binds to the C-terminal domain of the periplasmic host protein, TolA, presumably after pili retraction (21, 45, 99). The D1 and D2 domains are separated by a disordered, glycine-rich linker (45, 46, 71). This pathway of membrane penetration has been suggested to be similar for other filamentous phages (45). In the case of the filamentous phage f1, the TolR and TolQ proteins are also involved in the infection process (12, 13).

Phage particle assembly takes place at the periplasmic membrane in a highly ordered fashion. The single-stranded DNA coated with P5 protein is packaged "online" in the periplasmic (or outer) membrane by the addition of

* This gene-product nomenclature will be used for all filamentous phages.

Table 44-1 Current Phage Display Options

Phage	Fusion protein	Type of fusion	
		N-terminal	C-terminal
Filamentous	P3	Yes	Yes
	P8	Yes	–
	P6	–	Yes
	P7/P9	Yes	–
	Jun-Fos	–	Yes
Lambda	V	–	Yes
	D	Yes	Yes
T4	Wac	–	Yes
	Soc	Yes	Yes
	Hoc	Yes	–
T7	gp10	–	Yes

processed P8 protein residing in the membrane while P5 protein is removed. A normal assembly process results in filamentous phage being released from the infected cells without effecting cell lysis but instead maintains viability and colony-forming ability.

The P3 protein can be divided into two functional parts: the N-terminus required for infectivity and the C-terminus (amino acids 198–406) for proper particle morphogenesis (18). During morphogenesis the P3 protein (embedded in the membrane) plays an important role by terminating the assembly and protecting the P6 protein from degradation (95). Furthermore, the C-terminal domain of protein P3 is critical for membrane release as well as for virion stability (93). Particles devoid of P3 show aberrant sizes and are obviously noninfectious. Until recently, P3-based protein display has been achieved by N-terminal fusions leaving the C-terminal part of gene *3* intact, and it was thought that this part could not be engineered for display.

Fuh and Sidhu (37) have subsequently shown, however, that the C-terminus of the P3 protein can be manipulated to allow display in a M13 phagemid system. In this case, C-terminal fusions were achieved via optimized linker sequences. This type of display is suitable for functional complementary DNA (cDNA) cloning since it avoids the stop codon problem and also for the study of protein–protein interactions requiring free C-termini. Prior to the discovery of C-terminal domain display by protein P3 in filamentous phage, Crameri and Suter (17) had designed a two-component vector system suitable for cDNA display. They took advantage of the leucine zipper set, Jun and Fos, as a way to forge P3 and the expressed cDNA protein together using disulfide bonds located in the zipper. This system has been used to isolate allergens from of a number of sources via cDNA display (16).

Proteins P3- and P8-based systems differ in their capacity to display proteins. P3-based systems are able to display both peptides and functional proteins whereas display by P8, the major coat protein, is restricted to small peptides, such as a 6-mer. There are a number of reasons for this restriction: (i) displays of larger peptide reduces phage viability, for example 16 amino acids displayed via the P8 protein reduces phage viability to 1% (51, 72, 73); and (ii) the first five amino acids of the P8 N-terminus can vary but their deletion also results in reduced phage viability, though insertion of a pentapeptide between position 4 and 5 did not affect the helical symmetry of the phage and the peptide was exposed in an extended form (62). On the other hand, proline-rich sequences with large hydrophobic residues at position 7 and Asn at position 1 in the P8::peptide fusions were found to enhance particle viability (50). The amino acid residues present in the peptides are normally accessible except for those located 47 residues or fewer from the C-terminus of P8 (119).

The filamentous phage display vectors, including those of the phagemid types, are derived from phages fd, f1, M13 or modular hybrids of theirs. Vector options for P3 and P8 display can be described as type 3, 3+3, and 33 as well as 8, 8+8, and 88 (114). Phagemids are used in 3+3 or 8+8 display where wild-type P3 or P8 respectively are supplied in *trans* by a helper phage. In type 33 or 88 display, wild-type and the recombinant-gene version reside in the same phage chromosome. A genetically stable fd type 88 peptide display vector (Fth1), giving rise to high titers of recombinant phages, has recently been described (34).

P8::protein fusions allow less than 1 copy displayed per particle, but Nakayama et al. (84) were able to improve this type of display 10-fold via mutational "tricks." Recently, mutations in gene *8* have been reported to improve up to 100-fold the display of large proteins including oligomeric protein (110). Likewise, display of the Stoffel fragment of Taq polymerase as a fusion with P3 protein was improved more than 50-fold by selection of mutations of the signal sequence originating from the *pelB*-encoding pectate lyase of *Erwinia carotovora* that is often present in phagemid constructs (56). But this is a case-by-case approach since the mutations selected are not expected to support the same increase in display for every protein.

In a commonly used phagemid-based system, dependent on a helper phage to supply wild-type P3, only a small fraction of the total phage particles show display after assembly. To improve the frequency of display, gene *3*-deleted helper phage have been used in combination with plasmids providing wild-type P3 (27, 94). The recently introduced M13 helper phage ("hyperphage") appears to be a further improvement of this approach. A 400-fold increase in the display of single-chain antibodies has been reported using the "hyperphage" (100). In this case there is no competition by wild-type P3 during the assembly of the virions and the full copy number of P3 display is achieved. This approach, however, leads to loss of sometimes useful monovalent display as there are no wild-type P3 present to dilute the P3::protein fusion.

Display of fused proteins by the P6 protein is now possible (55). This system has been proven useful for cDNA display as the fusion takes place at the C-terminus of P6 which thereby alleviates the stop-codon problem experienced by the N-terminal display (1, 36, 49, 111, 123). More recently, the proteins P7 and P9 of phage M13 have been shown to work in functional co-display of antibody heavy- and light-chain variable regions in a phagemid format (38). In this case the chains were fused to the N-termini of the P7 and P9 proteins, respectively, and found to interact to form a functional Fv-binding domain on the phage surface. The engineering of filamentous phage M13 for phage display has recently been reviewed (109).

Genomic DNA-derived filamentous phage display libraries have proven very useful for identification of bacterial proteins as well as domains that interact with a range of target molecules (23, 54, 102). Indeed, a novel IgG-binding protein has been identified in *Staphylococcus aureus*, thereby demonstrating the power of this type of display (131). Product toxicity or failure of filamentous phage release in turn may hamper the display analysis. Therefore, special phagemid systems designed to facilitate toxic-product display have been constructed (7). However, the emerging display systems based on tailed phages such as λ, T4, and T7, which are released by cell lysis, offers a novel window of display by avoiding translocation of the protein fusions across the periplasmic membrane.

Phage λ Display

In phage λ, the major capsid protein, E, makes up an icosahedral capsid of 415 monomers. A tail connector is positioned at one of the capsid vertices. The tail is made up of 192 copies of the *V* gene product. However, λ lacked tail fibers until Hendrix and Duda (43) showed that λ *PaPa*, used in all λ studies, was a fiberless mutant (for a review of phage λ, see chapter 27). It was then envisioned that λ and its tail fibers could be used as a potential display system. The first report describing λ display, however, used its tail protein V for display (74). λ*foo* was designed for display at the C-terminal end of V. Even though it suffered from poor display efficiency (a few molecules per particle) and low phage yields, display of a homo-multimeric protein such as β-galactosidase was achieved. λ*foo* has also been used in epitope mapping (65, 66, 79).

Dunn (28) developed a C-terminal λ V-display system for presentation of a peptide (RRAVS) as target for cAMP-dependent protein kinase and in a subsequent report it was shown that the V-displaying peptide could complement a V-defective mutant phage to essentially normal phage yields (29). As for λ*foo*, only a limited display of β-galactosidase was observed using Dunn's V-display system. In a subsequent study a limited display of the α-complementation peptide of the β-galactosidase system was achieved. Such purified α-peptide phages also functioned in an in vitro α-complementation assay of β-galactosidase (30). λV display of a RDG sequence made the phages able to transfect COS cells at a significant frequency (31).

Several phages are known to strengthen their capsids after assembly by the addition of special phage-encoded decoration proteins. In the phage λ capsid, gpD of 11 kDa is present in 405 copies. Deposition of 135 trimeric D units on its capsid surface (26) is essential for stable head formation but certain chromosomal deletions can compensate for absence of D. A crystal-structure analysis of D in combination with cryo-electron microscopy and image reconstruction reveal that its N-terminus is disordered up to Ser 15 whereas the C-terminus is well ordered. Both termini are positioned on the same side of the trimer that binds to the capsid (130). Despite this seemingly awkward orientation, D works as a display platform for fusion proteins connected to it by linker peptides at either termini, as demonstrated by Sternberg and Hoess (116) and Mikawa et al. (76). Work in progress also shows that a 15 amino acid peptide of the N-terminal sequence of D is able to bind the expanded λ capsid (K.A. Miroshnikor et al., personal communication, 77).

Sternberg and Hoess (116) first constructed display plasmids in which D-fusion expression was under control of the *pTrc* or the more effectively regulated *ara* promoter. By utilizing the *cre-lox* recombination system of phage P1, the gpD-expressing plasmid with *loxP* can be picked up by an infecting D^- λ chromosome containing a *loxP* site in cells expressing Cre recombinase (see chapter 24 for a review of phage P1). λ particles, with the inserted displaying plasmid, were recovered as ampicillin-resistant transductants/lysogens. The lysogens were then induced to yield a library of λ particles that displayed peptides or proteins as N-terminal D fusions (8, 116). This λ display system has been further engineered for C-terminal D display of cDNA (104). A certain instability, however, of the λ plasmid cointegrates — even in the absence of Cre — has prompted the construction of C-terminal D display vectors, λ *171LoxP*$^-$ (105) and λ *Dsplay1* (10. 133). Both these vectors have been used for construction and panning of cDNA display libraries and λ *Dsplay1* was also compared with T7 cDNA display (133).

λ*foo* has also been used for engineering display at the N- or C-terminus of D (76). Plasmid vectors were designed for insertions in an engineered site between the first and the second codon of D, respectively. Plasmids for C-terminal fusion were also constructed by allowing insertion immediately after the termination codon of the *D* gene. The termination codon was then used to regulate the number of fusion proteins per phage particle by conditional chain terminations using *E. coli* suppressors. λ mutants lacking the D protein could then be complemented by these high copy D-fusion-expressing plasmid leading to particle display. Two phage vectors were also constructed for display at both termini of gpD: λ*fooDn* (N-terminal) and λ*fooDc* (C-terminal). The latter vector has been used for C-terminal display in a few studies (117, 118, 132) as well as a λ*foo* derivative called λ*fooDcSfi1* for C-terminal λ display of a cDNA library (87).

Phage T4 Display

T4 encodes two dispensable structural proteins, Soc and Hoc, which bind to the outer surface of mature T4 capsid (52). Soc is a 9 kDa protein and Hoc has a molecular weight of 40 kDa. Structural analysis including recent cryo-electron microscopy image reconstructions (53, 90) places Soc as a trimer at the triagonal points of the P23* lattice (the processed major capsid protein) whereas a single Hoc molecule is found at the center of each P23* hexamer. It is proposed that Soc functions like a clamp and protects the capsid against harsh conditions such as high pH and temperatures (see chapter 18 for a review of T4 biology). Both Soc and Hoc have been developed as display platforms. The first report to demonstrate surface display in T4, however, used the fibritin proteins encoded by the T4 gene *wac* (whisker's antigen control). Even though the C-terminus is essential for correct trimerization and folding of the fibritin protein (67), it could be extended by a 53 amino acid fusion to obtain functional T4 display (33).

There are 960 and 160 copies per particle of Soc and Hoc, respectively (57). Hence, Soc allows extensive multidisplay by T4 that can be extended by the use of polyheads (98). The rationale and operation of the display system developed by Ren et al. (98) are similar to those of Sternberg and Hoes's system (116). Namely, Soc and its fusion derivatives are expressed in *E. coli*, purified, and then bound in vitro to polyheads or in vivo using positive-selection vectors that force integration of *soc-fusion* sequences into a *soc*-deleted T4 chromosome. In this case the progeny phage carried, among others, a 312 amino acid sequence of poliovirus capsid protein, VP1, fused to the C-terminus of Soc. N-terminal peptide display using Hoc was achieved as part of an elegant procedure to clone linear DNA fragments in vivo. This work was later extended to a 183 amino acid functional N-terminal fusion of Hoc (96, 97). Jiang et al. (57) have also developed a T4 display system based on Soc as well as Hoc. In this case the N-termini were used as fusion points of a 36 amino acid PorA peptide from *Neisseria meningitidis*. The T4 display system appears little used so far.

Black and coworkers have described the development of an internal T4 display system based on the nonessential scaffolding protein IPIII that allows the construction, packaging, and even specific processing of proteins within the T4 capsid (47, 80, 81, 82). This expression–packaging–processing system (EPP) for internal display channels IPIII protein fusions into the T4 capsid and offers many applications such as detection and purification of proteins free from proteolysis as well as delivery of proteins into *E. coli* cells, and perhaps others.

Phage T7 Display

Phage T7 is released by cell lysis and the translocation of any fusion protein through the cell membrane/wall is avoided. T7 is a robust phage with a short latent period that should speed up the selection process. As in the case of phages λ and T4, the biology of T7 is very well understood (see chapter 20) and offers a variety of host-vector systems for a range of applications. The T7 capsid is composed of 415 copies of the capsid protein: 10 that are assembled into 60 hexamers on the faces of the icosahedron structure plus 11 pentamers at the head vertices. A short tail (genes *11* and *12*) with its six tail fibers (gene *17*) is attached at the remaining vertex through the head–tail connector (gene *8*). The capsid protein is made in two forms: 10A (344 amino acids) and 10B (397 amino acids). The production of 10B results from a translational frameshift at amino acid 341 of 10A and there is 10% of 10B in the capsid

The fact that functional T7 particles can be made from either of the capsid protein variants, 10A or 10B, including a range of mixtures of the two different protomers, prompted the exploration of the T7 capsid protein as a C-terminal display platform (101). The vector T7Select415-1b is reported to accept peptides up to 40–50 amino acids in length for display of 415 copies per particle. There are an increasing number of reports in which the T7Select415 system has been used (11, 48, 115).

T7Select1-1b is designed for display of peptides and proteins of less than 1 copy per phage particle and can tolerate fusions up to 900–1200 amino acids at that display density. In order to achieve the low-copy-number display (0.1–1 fusion per phage), the Phi10 promoter was removed and the original translation initiation site altered. Hence, other upstream promoters take over but at a reduced efficiency. Furthermore, a special 10A complementing host is needed to achieve the low-copy display. Construction of libraries in the T7Select systems utilizes vector arms and T7 packaging extracts. The T7Select1-1b system or its derivatives have been used for the display and panning of cDNA libraries against different targets including small-molecule chemical probes (42, 60, 85, 106, 107, 129). T7 display (Select1-2 series) has also been used to select RNA-binding proteins from cDNA libraries (20). In this case, the speed at which cycles of panning can be performed by the T7 system was mentioned explicitly.

Miscellaneous Phage (and Virus) Display

The tailspike protein (TSP) of phage P22 (phage P22 is reviewed in chapter 29), six copies of which are attached to the capsid to form the tail, has been used in peptide presentation (9). A 13 amino acid antigenic peptide from the VP1 protein of foot-and-mouth disease virus was joined to the C-terminal end of TSP under control of a *trp* promoter. Both the endorhamnosidase and assembly activities of the TSP- fusion were retained, indicating that P22-TSP system can be used for display.

Phage P4 (reviewed in chapter 26) encodes the capsid decorating protein Psu, which has been used for peptide presentation (68). Psu resides at each hexamer, probably as a dimer, thereby stabilizing the mature P4 capsid similar to D in phage λ and Soc in phage T4 (25). In Q_β (reviewed in chapter 15), functional virion display has been reported within the 195 amino acid extension of the coat protein A1 (63, 122) and recently, guided by the MS2 atomic structure, peptide display on live MS2 (also reviewed in chapter 15) has been achieved (121). Addition of tri- or hexapeptides at residue 100 of poliovirus caspsid protein, VP1, represents an implicit early example of virus display (14) and phage display has now been extended to both animal and plant virus systems (40).

Concluding Remarks

Based on current experience, most if not all phage/viral systems can probably be developed for display of peptides/proteins. The viability of the engineered particles sets the limits of display and in that respect the filamentous phages have proven very useful. Results on filamentous phage viability in organic solvent–water mixtures open up the possibility of employing phage display in nonaqueous media (89).

Although phage peptide display has been very successful in identifying ligands and epitopes (24, 61, 70) its use can sometimes be capricious (83, 120). It remains a challenge to assemble cDNA libraries for phage surface display that fully represent the cellular messengerRNA status. The recent engineering of C-terminal presentation in filamentous phage and the tailed phage systems is widening the window of cDNA display. Phage display is now an established technology as witnessed by an increasing number of applications, and more innovations are clearly due. A perhaps unexpected such innovation was the isolation of peptides with semiconductor binding specificities (124). These peptides have been reported to distinguish different crystallographic planes of gallium arsenide and silicon. Filamentous phage display was recently used to select peptides that favored top-phase partitioning of phage particles in a PEG/sodium phosphate two-phase system (3). In biosynthetic phage display, chemical synthesis is combined with the genetic diversity (32).

In recent years cell-free display systems, such as ribosome (41) and messengerRNA display (126), have been developed for making molecular libraries in vitro. These approaches are dependent on the *cis*-capture of the displayed protein by its own template, but relieved from cell-based transformation. Although procedures exist for generating high-complexity phage display peptide libraries ($\geq 10^9$) (88), the in vitro approach should give rise to even greater molecular/chemical diversities (35, 69), thereby improving the chances of enhancing weak affinities of molecular interactions (39). Phage display systems including certain libraries are now commercially available, and in a review on phage display and the development of tumor targeting agents Nilsson et al. (86) provide some comments on the intellectual property status of phage display.

Acknowledgment

This chapter was completed in January 2002. The author wishes to thank L. Black and M. Feiss for useful and inspiring comments and K. A. Miroshnikov, M. E. Cerritelli, G. Campusano, N. Cheng, J. F. Conway, and A. C. Steven for permission to mention their unpublished results.

References

1. Amery, L., G. P. Mannaerts, S. Subramani, P. P. Van Veldhoven, and M. Fransen. 2001. Identification of a novel human peroxisomal 2,4-dienoyl-CoA reductase related protein using the M13 phage protein VI phage display technology. Comb. Chem. High. Throughput Screen. 4:545–552.
2. Arap,W., R. Pasqualini, and E. Ruoslahti. 1998. Cancer treatment by targeted drug delivery to tumor vasculature in a mouse model. Science 279:377–380.
3. Bandmann, N., J. Van Alstine, A. Veide, and P. A. Nygren. 2002. Functional selection of phage displayed peptides for facilitated design of fusion tags improving aqueous two-phase partitioning of recombinant proteins. J. Biotechnol. 93:1–14.
4. Barbas, C. F., III, D. R. Burton, J. K. Scott, and G. J. Silverman(eds.). 2001. Phage Display: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
5. Barbas, C. F., III, A. S. Kang, R. A. Lerner, and S. J. Benkovic. 1991. Assembly of combinatorial antibody libraries on phage surfaces: the gene *III* site. Proc. Natl. Acad. Sci. USA 15:7978–7982.
6. Bass, S., R. Greene, and J. A.Wells. 1990. Hormone phage: an enrichment method for variant proteins with altered binding properties. Proteins 8:309–314.
7. Beekwilder, J., J. Rakonjac, M. Jongsma, and D. Bosch. 1999. A phagemid vector using the *E. coli* phage shock promoter facilitates phage display of toxic proteins. Gene 228:23–31.

8. Beghetto, E., A. Pucci, O. Minenkova, A. Spadoni, L. Bruno, W. Buffolano, D. Soldati, F. Felici, and N. Gargano. 2001. Identification of a human immunodominant B-cell epitope within the GRA1 antigen of *Toxoplasma gondii* by phage display of cDNA libraries. Int. J. Parasitol. 31:1659–1668.
9. Carbonell, X., and A. Villaverde. 1998. Insertional mutagenesis in the tailspike protein of bacteriophage P22. Biochem. Biophys. Res. Commun. 244:428–433.
10. Castagnoli, L., A. Zucconi, M. Quondam, M. Rossi, P. Vaccaro, S. Panni, S. Paoluzi, E. Santonico, L. Dente, and G. Cesareni. 2001. Alternative bacteriophage display systems. Comb. Chem. High Throughput Screen. 4:121–133.
11. Castillo, J., B. Goodson, and J. Winter. 2001. T7 displayed peptides as targets for selecting peptide specific scFvs from M13 scFv display libraries. J. Immunol. Methods 257:117–122
12. Click, E. M., and R. E. Webster. 1997. Filamentous phage infection: required interactions with the TolA protein, J. Bacteriol. 179:6464–6471.
13. Click, E. M., and R. E. Webster. 1998. The TolQRA proteins are required for membrane insertion of the major capsid protein of the filamentous phage f1 during infection. J. Bacteriol. 180:1723–1728.
14. Colbere-Garapin, F., C. Christodoulou, R. Crainic, A. C. Garapin, and A. Candrea. 1988. Addition of a foreign oligopeptide to the major capsid protein of poliovirus. Proc. Natl. Acad. Sci. USA 85:8668–8672
15. Cortese, R., F. Felici, G. Galfre, A. Luzzago, P. Monaci, and A. Nicosia. 1994. Epitope discovery using peptide libraries displayed on phage. Trends Biotechnol. 12:262–267.
16. Crameri, R., and R. Kodzius. 2001. The powerful combination of phage surface display of cDNA libraries and high throughput screening. Comb. Chem. High Throughput Screen. 4:145–155.
17. Crameri, R., and M. Suter. 1993. Display of biologically active proteins on the surface of filamentous phages: a cDNA cloning system for selection of functional gene products linked to the genetic information responsible for their production. Gene 137:69–75. (Erratum in Gene 1995; 160:139.)
18. Crissman, J. W., and G. P. Smith. 1984. Gene-*III* protein of filamentous phages: evidence for a carboxyl-terminal domain with a role in morphogenesis. Virology 132: 445–455.
19. Cwirla, S. E., E. A. Peters, R. W. Barrett, and W. J. Dower. 1990. Peptides on phage: a vast library of peptides for identifying ligands. Proc. Natl. Acad. Sci. USA 87:6378–6382.
20. Danner, S., and J. G. Belasco. 2001. T7 phage display: a novel genetic selection system for cloning RNA-binding proteins from cDNA libraries. Proc. Natl. Acad. Sci. USA 98:12954–12959 .
21. Deng, L. W., P. Malik, and R. N. Perham. 1999. Interaction of the globular domains of pIII protein of filamentous bacteriophage fd with the F-pilus of *Escherichia coli*. Virology 253:271–277.
22. Devlin, J. J., L. C. Panganiban, and P. E. Devlin. 1990. Random peptide libraries: a source of specific protein binding molecules. Science 249: 404–406.
23. Djojonegoro, B. M., M. J. Benedik, and R. C. Willson. 1994. Bacteriophage surface display of an immunoglobulin-binding domain of *Staphylococcus aureus* protein A. Biotechnology (N.Y.) 12:169–172.
24. D'Mello, F., and C. R. Howard. 2001. An improved selection procedure for the screening of phage display peptide libraries. J. Immunol. Methods 247:191–203.
25. Dokland, T., M. L. Isaksen, S. D. Fuller, and B. H. Lindqvist. 1993. Capsid localization of the bacteriophage P4 Psu protein. Virology 194:682–687.
26. Dokland, T., and H. Murialdo. 1993. Structural transitions during maturation of bacteriophage lambda capsids. J. Mol. Biol. 233:682–694.
27. Duenas, M., and C. A. Borrebaeck. 1995. Novel helper phage design: intergenic region affects the assembly of bacteriophages and the size of antibody libraries. FEMS Microbiol Lett. 125:317–321.
28. Dunn, I. S. 1995. Assembly of functional bacteriophage lambda virions incorporating C-terminal peptide or protein fusions with the major tail protein. J. Mol. Biol. 248:497–506.
29. Dunn, I. S. 1996. Total modification of the bacteriophage lambda tail tube major subunit protein with foreign peptides. Gene 183:15–21.
30. Dunn, I. S. 1996. In vitro alpha-complementation of beta-galactosidase on a bacteriophage surface. Eur. J. Biochem. 242:720–726.
31. Dunn, I. S. 1996. Mammalian cell binding and transfection mediated by surface-modified bacteriophage lambda. Biochimie 78:856–861.
32. Dwyer, M. A., W. Lu, J. J. Dwyer, and A. A. Kossiakoff. 2000. Biosynthetic phage display: a novel protein engineering tool combining chemical and genetic diversity. Chem. Biol. 7:263–274.
33. Efimov, V. P., I. V. Nepluev, and V. V. Mesyanzhinov. 1995. Bacteriophage T4 as a surface display vector virus. Genes 10:173–177.
34. Enshell-Seijffers, D., L. Smelyanski, and J. M. Gershoni. 2001. The rational design of a "type 88" genetically stable peptide display vector in the filamentous bacteriophage fd. Nucleic Acids Res. 29:E50–0.
35. FitzGerald, K. 2000. In vitro display technologies:new tools for drug discovery. Drug Discov. Today 5:253–258.
36. Fransen, M., P. P. Van Veldhoven, and S. Subramani. 1999. Identification of peroxisomal proteins by using M13 phage protein VI phage display: molecular evidence that mammalian peroxisomes contain a 2,4-dienoyl-CoA reductase. Biochem. J. 340:561–568.
37. Fuh, G., and S. S. Sidhu. 2000. Efficient phage display of polypeptides fused to the carboxy-terminus of the M13 gene-*3* minor coat protein. FEBS Lett. 480:231–234.
38. Gao, C., S. Mao, C. H. Lo, P. Wirsching, R. A. Lerner, and K. D. Janda. 1999. Making artificial antibodies: a format for phage display of combinatorial heterodimeric arrays. Proc. Natl. Acad. Sci. USA 96:6025–6030.

39. Gold, L. 2001. mRNA display: diversity matters during in vitro selection. Proc. Natl. Acad. Sci. USA 98:4825–4826.
40. Grabherr, R., and W. Ernst. 2001. The Baculovirus expression system as a tool for generating diversity by viral surface display. Comb. Chem. High Throughput Screen. 4:185–192.
41. Hanes, J., L. Jermutus, and A. Pluckthun. 2000. Selecting and evolving functional proteins in vitro by ribosome display. Methods Enzymol. 328:404–430.
42. Hansen, M. H., B. Ostenstad, and M. Sioud. 2001. Identification of immunogenic antigens using a phage-displayed cDNA library from an invasive ductal breast carcinoma tumour. Int. J. Oncol. 19:1303–1309.
43. Hendrix, R. W., and R. L. Duda. 1992. Bacteriophage lambda PaPa: not the mother of all lambda phages. Science 258:1145–1148.
44. Hoess, R. H. 2001. Protein design and phage display. Chem. Rev. 101:3205–3218.
45. Holliger P., and L. Riechmann. 1997. A conserved infection pathway for filamentous bacteriophages is suggested by the structure of the membrane penetration domain of the minor coat protein g3p from phage fd. Structure 5:265–275.
46. Holliger, P., L. Riechmann, and R. L. Williams. 1999. Crystal structure of the two N-terminal domains of g3p from filamentous phage fd at 1.9Å: evidence for conformational lability. J. Mol. Biol. 288:649–657.
47. Hong, Y. R., and L. W. Black. 1993. An expression-packaging-processing vector which selects and maintains 7-kb DNA inserts in the blue T4 phage genome. Gene 136:193–198.
48. Houshmand, H., G. Froman, and G. Magnusson. 1999. Use of bacteriophage T7 displayed peptides for determination of monoclonal antibody specificity and biosensor analysis of the binding reaction. Anal Biochem. 268:363–370.
49. Hufton, S. E., P. T. Moerkerk, E. V. Meulemans, A. de Bruine, J. W. Arends, and H. R. Hoogenboom. 1999. Phage display of cDNA repertoires: the pVI display system and its applications for the selection of immunogenic ligands. J. Immunol. Methods 231:39–51.
50. Iannolo, G., O. Minenkova, S. Gonfloni, L. Castagnoli, and G. Cesareni. 1997. Construction, exploitation and evolution of a new peptide library displayed at high density by fusion to the major coat protein of filamentous phage. Biol. Chem. 378:517–521.
51. Iannolo, G., O. Minenkova, R. Petruzzelli, and G. Cesareni. 1995. Modifying filamentous phage capsid: limits in the size of the major capsid protein. J. Mol. Biol. 248:835–844.
52. Ishii, T., and M. Yanagida. 1977. The two dispensable structural proteins (Soc and Hoc) of the T4 phage capsid; their purification and properties, isolation and characterization of the defective mutants, and their binding with the defective heads in vitro. J. Mol. Biol. 109:487–514.
53. Iwasaki, K., B. L. Trus, P. T. Wingfield, N. Cheng, G. Campusano, V. B. Rao, and A. C. Steven. 2000. Molecular architecture of bacteriophage T4 capsid: vertex structure and bimodal binding of the stabilizing accessory protein, Soc. Virology 271:321–333.
54. Jacobsen, K., and L. Frykberg. 2001. Shotgun phage display cloning. Comb. Chem. High Throughput. Screen. 4:135–143.
55. Jespers, L. S., J. H. Messens, A. De Keyser, D. Eeckhout, I. Van den Brande, Y. G. Gansemans, M. J. Lauwereys, G. P. Vlasuk, and P. E. Stanssens. 1995. Surface expression and ligand-based selection of cDNAs fused to filamentous phage gene VI. Biotechnology (N.Y.) 13:378–382.
56. Jestin, J. L., G. Volioti, and G. Winter. 2001. Improving the display of proteins on filamentous phage. Res. Microbiol. 152:187–191.
57. Jiang, J., L. Abu-Shilbayeh, and V. B. Rao. 1997. Display of a PorA peptide from *Neisseria meningitidis* on the bacteriophage T4 capsid surface. Infect. Immun. 65:4770–4777.
58. Johnsson, K., and L. Ge. 1999. Phage display of combinatorial peptide and protein libraries and their applications in biology and chemistry. Curr. Top. Microbiol. Immunol. 243:87–105.
59. Kang, A. S., C. F. Barbas, K. D. Janda, S. J. Benkovic, and R. A. Lerner. 1991. Linkage of recognition and replication functions by assembling combinatorial antibody Fab libraries along phage surfaces. Proc. Natl. Acad. Sci. USA 88:4363–4366.
60. Kataoka, K., K. Yoshitomo-Nakagawa, S. Shioda, and M. Nishizawa. 2001. A set of Hox proteins interact with the Maf oncoprotein to inhibit its DNA binding, transactivation, and transforming activities. J. Biol. Chem. 276:819–826.
61. Kay, B. K., J. Kasanov, and M. Yamabhai. 2001. Screening phage-displayed combinatorial peptide libraries. Methods 24:240–246.
62. Kishchenko, G., H. Batliwala, and L. Makowski. 1994. Structure of a foreign peptide displayed on the surface of bacteriophage M13. J. Mol. Biol. 241:208–213.
63. Kozlovska, T. M., I. Cielens, I. Vasiljeva, A. Strelnikova, A. Kazaks, A. Dislers, D. Dreilina, V. Ose, I. Gusars, and P. Pumpens. 1996. RNA phage Q beta coat protein as a carrier for foreign epitopes. Intervirology 39:9–15.
64. Krebs, B., R. Rauchenberger, S. Reiffert, C. Rothe, M. Tesar, E. Thomassen, M. Cao, T. Dreier, D. Fischer, A. Hoss, L. Inge, A. Knappik, M. Marget, P. Pack, X. Q. Meng, R. Schier, P. Sohlemann, J. Winter, J. Wolle, and T. Kretzschmar. 2001. High-throughput generation and engineering of recombinant human antibodies. J. Immunol. Methods 254:67–84.
65. Kuwabara, I., H. Maruyama, S. Kamisue, M. Shima, A. Yoshioka, and I. N. Maruyama. 1999. Mapping of the minimal domain encoding a conformational epitope by lambda phage surface display: factor VIII inhibitor antibodies from haemophilia A patients. J. Immunol. Methods 224:89–99.
66. Kuwabara, I., H. Maruyama, Y. G. Mikawa, R. I. Zuberi, F. T. Liu, and I. N. Maruyama. 1997. Efficient epitope mapping by bacteriophage lambda surface display. Nat. Biotechnol. 15:74–78.
67. Letarov, A. V., Y. Y. Londer, S. P. Boudko, and V. V. Mesyanzhinov. 1999. The carboxy-terminal domain

initiates trimerization of bacteriophage T4 fibritin. Biochemistry (Mosc.) 64:817–823.

68. Lindqvist, B. H., and S. Naderi. 1995. Peptide presentation by bacteriophage P4. FEMS Microbiol. Rev. 17:33–39.
69. Lohse, P., and M. C. Wright. 2001. In vitro protein display in drug discovery. Curr. Opin. Drug Discov. Dev. 4:198–204.
70. Lowman, H. B. 1997. Bacteriophage display and discovery of peptide leads for drug development. Annu. Rev. Biophys. Biomol. Struct. 26:401–424.
71. Lubkowski, J., F. Hennecke, A. Pluckthun, and A. Wlodawer. 1998. The structural basis of phage display elucidated by the crystal structure of the N-terminal domains of g3p. Nat. Struct. Biol. 5:140–147.
72. Malik, P., T. D. Terry, F. Bellintani, and R. N. Perham. 1998. Factors limiting display of foreign peptides on the major coat protein of filamentous bacteriophage capsids and a potential role for leader peptidase. FEBS Lett. 436:263–266.
73. Malik, P., T. D. Terry, L. R. Gowda, A. Langara, S. A. Petukhov, M. F. Symmons, L. C. Welsh, D. A. Marvin, and R. N. Perham. 1996. Role of capsid structure and membrane protein processing in determining the size and copy number of peptides displayed on the major coat protein of filamentous bacteriophage. J. Mol. Biol. 260:9–21.
74. Maruyama, I. N., H. I. Maruyama, and S. Brenner. 1994. Lambda *foo*: a lambda phage vector for the expression of foreign proteins. Proc. Natl. Acad. Sci. USA 91:8273–8277.
75. McCafferty, J., A. D. Griffiths, G. Winter, and D. J. Chiswell. 1990. Phage antibodies: filamentous phage displaying antibody variable domains. Nature 348:552–554.
76. Mikawa, Y. G., I. N. Maruyama, and S. Brenner. 1996. Surface display of proteins on bacteriophage lambda heads. J. Mol. Biol. 262:21–30.
77. Miroshnikov, K. A., E. M. Cerritelli, G. Campusano, N. Cheng, F. J. Conway, and C. A. Steven. Personal communication.
78. Monaci, P., L. Urbanelli, and L. Fontana. 2001. Phage as gene delivery vectors. Curr. Opin. Mol. Ther. 3:159–169.
79. Moriki, T., I. Kuwabara, F. T. Liu, and I. N. Maruyama. 1999. Protein domain mapping by lambda phage display: the minimal lactose-binding domain of galectin-3. Biochem. Biophys. Res. Commun. 265:291–296.
80. Mullaney, J. M., and L. W. Black. 1996. Capsid targeting sequence targets foreign proteins into bacteriophage T4 and permits proteolytic processing. J. Mol. Biol. 261:372–378.
81. Mullaney, J. M. and L. W. Black. 1998. Activity of foreign proteins targeted within the bacteriophage T4 head and prohead: implications for packaged DNA structure. J. Mol. Biol. 283:913–929.
82. Mullaney, J. M., and L. W. Black. 1998. GFP:HIV-1 protease production and packaging with a T4 phage expression–packaging–processing system. Biotechniques 25:1008–1012.
83. Murthy, K. K., I. Ekiel, S. H. Shen, and D. Banville. 1999. Fusion proteins could generate false positives in peptide phage display. Biotechniques 26:142–149.
84. Nakayama, G. R., G. Valkirs, D. McGrath, and W. D. Huse. 1996. Improving the copy numbers of antibody fragments expressed on the major coat protein of bacteriophage M13. Immunotechnology 2:197–207.
85. Nakielny, S., S. Shaikh, B. Burke, and G. Dreyfuss. 1999. Nup153 is an M9-containing mobile nucleoporin with a novel Ran-binding domain. EMBO J. 18:1982–1995.
86. Nilsson, F., L. Tarli, F. Viti, and D. Neri. 2000. The use of phage display for the development of tumour targeting agents. Adv. Drug Deliv. Rev. 43:165–196.
87. Niwa, M., H. Maruyama, T. Fujimoto, K. Dohi, and I. N. Maruyama. 2000. Affinity selection of cDNA libraries by lambda phage surface display. Gene 256: 229–236.
88. Noren, K. A., and C. J. Noren. 2001. Construction of high-complexity combinatorial phage display peptide libraries. Methods 23:169–178.
89. Olofsson, L., J. Ankarloo, P. O. Andersson, and I. A. Nicholls. 2001. Filamentous bacteriophage stability in non-aqueous media. Chem. Biol. 8:661–671.
90. Olson, N. H., M. Gingery, F. A. Eiserling, and T. S. Baker. 2001. The structure of isometric capsids of bacteriophage T4. Virology 279:385–391.
91. Pasqualini, R., and E. Ruoslahti. 1996. Organ targeting in vivo using phage display peptide libraries. Nature 380:364–366.
92. Poul, M. A., and J. D. Marks. 1999. Targeted gene delivery to mammalian cells by filamentous bacteriophage. J. Mol. Biol. 288:203–211.
93. Rakonjac, J., J. Feng and P. Model. 1999. Filamentous phage are released from the bacterial membrane by a two-step mechanism involving a short C-terminal fragment of pIII. J. Mol. Biol. 289: 1253–1265.
94. Rakonjac, J., G. Jovanovic, and P. Model. 1997. Filamentous phage infection-mediated gene expression: construction and propagation of the gIII deletion mutant helper phage R408d3. Gene 198:99–103.
95. Rakonjac, J., and P. Model. 1998. Roles of pIII in filamentous phage assembly. J. Mol. Biol. 282:25–41.
96. Ren, Z.J., R.G. Baumann, and L.W. Black. 1997. Cloning of linear DNAs in vivo by overexpressed T4 DNA ligase: construction of a T4 phage *hoc* gene display vector. Gene 195:303–311.
97. Ren, Z., and L. W. Black. 1998. Phage T4 SOC and HOC display of biologically active, full-length proteins on the viral capsid. Gene 215:439–444.
98. Ren, Z. J., G. K. Lewis, P. T. Wingfield, E. G. Locke, A. C. Steven, and L. W. Black. 1996. Phage display of intact domains at high copy number: a system based on SOC, the small outer capsid protein of bacteriophage T4. Protein Sci. 5:1833–1843.
99. Riechmann, L., and P. Holliger. 1997. The C-terminal domain of TolA is the coreceptor for filamentous phage infection of *E. coli*. Cell 90:351–360.
100. Rondot, S., J. Koch, F. Breitling, and S. Dubel. 2001. A helper phage to improve single-chain antibody presentation in phage display. Nat. Biotechnol. 19:75–78.
101. Rosenberg, A., K. Griffin, F. W. Studier, M. McCormick, J. Berg, R. Novy, and R. Mierendorf. 1996. T7 Select^R

phage display system: a powerful new protein display system based on bacteriophge T7. InNovations 6:1–6.

102. Rudgers, G.W., and T. Palzkill. 2001. Protein minimization by random fragmentation and selection. Protein Eng. 14:487–492.

103. Ruoslahti, E. 2000. Targeting tumor vasculature with homing peptides from phage display. Semin. Cancer Biol. 10:435–442.

104. Santi, E., S. Capone, C. Mennuni, A. Lahm, A. Tramontano, A. Luzzago, and A. Nicosia. 2000. Bacteriophage lambda display of complex cDNA libraries: a new approach to functional genomics. J. Mol. Biol. 296:497–508

105. Santini, C., D. Brennan, C. Mennuni, R. H. Hoess, A. Nicosia, R. Cortese, and A. Luzzago. 1998. Efficient display of an HCV cDNA expression library as C-terminal fusion to the capsid protein D of bacteriophage lambda. J. Mol. Biol. 282:125–135.

106. Savinov, S. N., and D. J. Austin. 2001. The cloning of human genes using cDNA phage display and small-molecule chemical probes. Comb. Chem. High Throughput Screen. 4:593–597 .

107. Sche, P. P., K. M. McKenzie, J. D. White, and D. J. Austin. 2001. Corrigendum to: "Display cloning: functional identification of natural product receptors using cDNA-phage display" [Chem. Biol. 6 (1999) 707–716]. Chem. Biol. 8:399–400.

108. Scott, J. K., and G. P. Smith. 1990. Searching for peptide ligands with an epitope library. Science 249:386–390.

109. Sidhu, S. S. 2001. Engineering M13 for phage display. Biomol. Eng. 18:57–63.

110. Sidhu, S. S., G.A. Weiss, and J.A. Wells. 2000. High copy display of large proteins on phage for functional selections. J. Mol. Biol. 296:487–495.

111. Sioud, M., and M. H. Hansen. 2001. Profiling the immune response in patients with breast cancer by phage-displayed cDNA libraries. Eur. J. Immunol. 31:716–725.

112. Sioud, M., M. Hansen, and A. Dybwad. 2000. Profiling the immune responses in patient sera with peptide and cDNA display libraries. Int. J. Mol. Med. 6:123–128.

113. Smith, G. P. 1985. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. Science 228:1315–1317.

114. Smith, G. P. 1993. Surface display and peptide libraries. Gene 128:1–2.

115. Sokoloff, A. V., I. Bock, G. Zhang, M. G. Sebestyen, and J. A. Wolff. 2000. The interactions of peptides with the innate immune system studied with use of T7 phage peptide display. Mol. Ther. 2:131–139.

116. Sternberg, N., and R. H. Hoess. 1995. Display of peptides and proteins on the surface of bacteriophage lambda. Proc. Natl. Acad. Sci. USA 92:1609–1613.

117. Stolz, J., A. Ludwig, and N. Sauer. 1998. Bacteriophage lambda surface display of a bacterial biotin acceptor domain reveals the minimal peptide size required for biotinylation. FEBS Lett. 440:213–217.

118. Stolz, J., A. Ludwig, R. Stadler, C. Biesgen, K. Hagemann, and N. Sauer. 1999. Structural analysis of a plant sucrose carrier using monoclonal antibodies and bacteriophage lambda surface display. FEBS Lett. 453:375–379.

119. Terry, T. D., P. Malik, and R. N. Perham. 1997. Accessibility of peptides displayed on filamentous bacteriophage virions: susceptibility to proteinases. Biol. Chem. 378:523–530.

120. Vanhoorelbeke, K., R. M. van der Plas, G. Vandecasteele, S. Vauterin, E. G. Huizinga, J. J. Sixma, and H. Deckmyn. 2000. Sequence alignment between vWF and peptides inhibiting the vWF-collagen interaction does not result in the identification of a collagen-binding site in vWF. Thromb. Haemost. 84:621–625.

121. van Meerten, D., R. C. Olsthoorn, J. van Duin, and R. M. Verhaert. 2001. Peptide display on live MS2 phage: restrictions at the RNA genome level. J. Gen. Virol. 82:1797–1805.

122. Vasiljeva I., T. Kozlovska, I. Cielens, A. Strelnikova, A. Kazaks, V. Ose, and P. Pumpens. 1998. Mosaic Qbeta coats as a new presentation model. FEBS Lett. 431:7–11.

123. Viaene, A., A. Crab, M. Meiring, D. Pritchard, and H. Deckmyn. 2001. Identification of a collagen-binding protein from *Necator americanus* by using a cDNA-expression phage display library. J. Parasitol. 87:619–625.

124. Whaley, S. R., D. S. English, E. L. Hu, P. F. Barbara, and A. M. Belcher. 2000. Selection of peptides with semiconductor binding specificity for directed nanocrystal assembly. Nature 405:665–668.

125. Wilson, D. R., and B. B. Finlay. 1998. Phage display: applications, innovations, and issues in phage and host biology. Can. J. Microbiol. 44:313–329.

126. Wilson, D. S., A. D. Keefe, and J.W. Szostak. 2001. The use of mRNA display to select high-affinity protein-binding peptides. Proc. Natl. Acad. Sci. USA 98:3750–3755.

127. Winthrop, M. D., G. L. Denardo, and S. J. Denardo. 2000. Antibody phage display applications for nuclear medicine imaging and therapy. Q. J. Nucl. Med. 44:284–295.

128. Wittrup, K. D. 1999. Phage on display. Trends Biotechnol. 17:423–424.

129. Yamamoto, M., Y. Kominato, and F. Yamamoto. 1999. Phage display cDNA cloning of protein with carbohydrate affinity. Biochem. Biophys. Res. Commun. 255:194–1999.

130. Yang, F., P. Forrer, Z. Dauter, J. F. Conway, N. Cheng, M. E. Cerritelli, A. C. Steven, A. Pluckthun, and A. Wlodawer. 2000. Novel fold and capsid-binding properties of the lambda-phage display platform protein gpD. Nat. Struct. Biol. 7:230–237.

131. Zhang, L., K. Jacobsson, J. Vasi, M. Lindberg, and L. Frykberg. 1998. A second IgG-binding protein in *Staphylococcus aureus*. Microbiology 144:985–991.

132. Zhang, Y., J. W. Pak, I. N. Maruyama, and M. Machida. 2000. Affinity selection of DNA-binding proteins displayed on bacteriophage lambda. J. Biochem. (Tokyo) 127:1057–1063.

133. Zucconi, A., L. Dente, E. Santonico, L. Castagnoli, and G. Cesareni. 2001. Selection of ligands by panning of domain libraries displayed on phage lambda reveals new potential partners of synaptojanin 1. J. Mol. Biol. 307:1329–1339.

45

Bacteriophage as Pollution Indicators

CHARLES P. GERBA

Because of the difficulty and cost of detecting waterborne enteric pathogens, indicator organisms have been used since the beginning of the twentieth century. Much of the work in this area has focused on bacterial indicators, such as coliform and fecal coliform bacteria. However, it has been recognized in the last 30 years that these traditional indicators do not always reflect the waterborne occurrence of human pathogenic viruses and protozoa. Thus, bacteriophages have been investigated as a better indicator of these groups of pathogens in water, as models of enteric virus removal by treatment processes, and to examine the fate and transport of enteric viruses in the environment (12).

The term "indicator organism" is often not clearly defined. By contrast, an "index organism" is usually defined as one related to the occurrence of a selected surrogate microorganism or microorganisms (table 45-1). The relationship may be direct, such as an index of human viruses, or indirect, such as an index of fecal pollution or types of fecal pollution (i.e., human or animal) (27). The criteria for an index organism are very similar to those commonly used for bacterial fecal indicators. An indicator organism, on the other hand, is measured to check the performance of a treatment process against previously set standards. For example, an indicator is used to evaluate the performance of drinking-water disinfection for the inactivation of enteric viruses. To serve as an effective indicator, the resistance of the indicator organism and the pathogen to the disinfectant should be similar.

Three main groups of bacteriophages infecting enteric bacteria have received the greatest amount of study in the assessment of water quality. These are the somatic coliphages, the F-specific RNA coliphages, and the bacteriophages infecting *Bacteroides fragilis* (table 45-2). This chapter largely concerns the use of these bacteriophages as indicators of fecal pollution and as index organisms of pathogenic human enteric viruses.

Somatic Coliphages

The use of bacteriophages as a test of fecal pollution was originally suggested by Coetzee (3) and Kott (31). Kott showed that *Escherichia coli* B phages were more resistant in the water environment than coliform bacteria because the ratio of phages to coliforms shifted from 1:100 in the vicinity of sewage outfalls to 1:1 to 1:10 at more distant locations. The group of bacteriophages infecting *E. coli* B and related host strains are now described as somatic coliphages. Natural host strains of somatic coliphages include *E. coli* and closely related bacterial species. Somatic coliphages are a heterogeneous group of phages belonging to the *Mycoviridae, Siphoviridae, Podovirdae,* and *Microviridae* families (see chapter 2 for a review of phage classification). The specific assay methods determine which phages are detected and which are not. In this respect, somatic coliphages are a method-defined parameter similar to the total and fecal coliform groups of indicator bacteria, with inherent disadvantages with regard to method standardization and comparison of results from different studies. In addition, somatic coliphages are a heterogeneous group with regard to their response to environmental factors (i.e., temperature, pH) and at least some members of the group are able to multiply in waters that are not subject to fecal contamination. However, the contribution of this potential replication outside the gut to their occurrence in natural environments has never been quantified (27). Bacteriophages frequently used to study somatic coliphage behavior are T-even and T-odd, ϕX174, and PRD1. The biology of these various phages are reviewed in chapters 18 (T-even), 17, 19, 20 (T-odd), 11 (ϕX174), and 13 (PRD1).

Coliphages infecting *E. coli* C are easily detected in the feces of man, cattle, pigs, chickens, and other animals (5, 16, 21). Phage numbers vary widely between <10 and 10^8 plaque forming units (PFU) per gram. The highest

Table 45-1 Definitions for Indicator and Index Microorganisms of Public Health Concern

Group	Definition
Process indicator	A group of organisms that demonstrates the efficacy of a process
Fecal indicator	A group of organisms that indicates the presence of fecal contamination. They only infer that pathogens may be present
Index and model organisms	A group/or species indicative of pathogen presence and behavior respectively, such as *E. coli* as an index for *Salmonella* and F-specific RNA bacteriophages as models of human enteric viruses

From Ashbolt et al. (1).

Table 45-2 Major Bacteriophage Groups Considered Appropriate Virus Models in the Environment

Description	Host Strain	Comments
Somatic coliphages	*E. coli* C (most commonly used)	Heterogeneous group of different morphology; frequent occurrence in human and animal feces (10^2–$10^8 g^{-1}$) and wastewater (10^3–10^4 ml^{-1}); may multiply in the environment; good persistence in the environment; readily inactivated by water treatment processes (with the exception of a few types)
F-specific RNA bacteriophages	*S. typhimurium* phage type 3 Nal^r (F = 42 lac::Tn5), *E. coli* HS[pFamp]R	Homogeneous group, physical properties similar to those of enteroviruses; infrequent in human and animal feces (up to 10^3 g^{-1}), frequent occurrence in wastewater (10^3–10^4 ml^{-1}); can multiply only at temperatures above 30°C; relatively high resistance; serotypes may be related to the (human or animal) origin of fecal pollution
B. fragilis phages	*B. fragilis* HSP40	Occur only in human feces (up to 10^8 g^{-1}); do not multiply in the environment; host strain possibly not applicable around the world; relatively low numbers in wastewater (<1–10^3 ml^{-1}); relatively homogeneous group; relatively high resistance

numbers have usually been found in the feces of pigs, calves, and boiler chickens. In human feces, somatic coliphages are lower in number (often <10 PFU/g) or undetectable. The arithmetic mean concentration of somatic coliphages varies from 10^4 to 10^7 per gram, which is three orders of magnitude lower than the fecal concentration of *E. coli* (10^7 to 10^9 per gram) (11). Somatic coliphages have been detected in 2.5–88% of samples of human feces in studies from Europe, United States, Asia, and South Africa (14). Relatively low isolation frequencies were found in studies in Japan, possibly related to the choice of a relatively insensitive host strain (14).

Somatic coliphages are the most abundant type found in raw, untreated domestic sewage, with values ranging from 10^4 to 10^5 PFU/ml (4, 8, 18). They are also found in similar numbers in abattoir wastewater and animal slurries (table 45-3).

Table 45-3 Typical Concentrations of Enteroviruses, Bacteriophages, and Fecal Bacteria in Domestic, Hospital, and Slaughterhouse Wastewater

Microorganism	Concentration (PFU or CFU/ml)
Enteroviruses	1×10^{-1}
Somatic coliphages	1×10^4
F-specific RNA phages	3×10^3
B. fragilis phages	1×10^2 $<10^{-2}$ in slaughterhouse wastewater
Fecal coliforms	1×10^5
Fecal streptococci	1×10^4

PFU, plaque forming units; CFU = colony forming units.

F-specific RNA Bacteriophages

F-specific RNA bacteriophages have simple polyhedral symmetry (icosahedron), are 21–30 nm in diameter, and contain single-stranded RNA as the genome. They belong to the family *Leviviridae*. They infect bacteria through the sex pili, which are coded by the F plasmid. The F plasmid is transferable to a wide range of Gram-negative bacteria. The pili encoded by the F plasmid do not form below 25°C (37). Therefore, the probability of F-specific phages replicating in the environment is low. The infection process is inhibited by the presence of RNase in the assay medium, which can be used to distinguish between the F-specific bacteriophages and the rod-shaped F-specific

DNA bacteriophages of the family *Inoviridae*, which also infect the host cell through the pili (27). Chapter 15 reviews the biology of single-stranded RNA viruses, and chapter 40 discusses the biology of *Inoviridae* phage of mycoplasma.

Salmonella typhimurium WG49 and *E. coli* HS strains have been modified to detect F-specific bacteriophages, but will also detect a small number of somatic phages (17) (table 45-4). All phages detected by the modified strains are usually referred to as F-specific bacteriophages. The number of F-specific RNA bacteriophages is the difference between the number of phages counted in the presence and in the absence of RNase in the assay medium. More than 90% of the phages detected in sewage by the modified strains are F-specific RNA bacteriophages (27).

F-specific RNA phages are less frequently found in animal feces than somatic phages. Dhillon et al. (5), using identification of plaques obtained from male *E. coli*, did not detect F-specific phages in feces from cows, pigs and humans. Osawa et al. (33) likewise did not detect F-specific RNA phages in birds, pigs, and cows. They were detected in 2% of the feces from humans and 2% of horse feces. Havelaar et al. (16, 21) used the host strain WG49 and found similar results: only the feces of boiler chickens were relatively constant sources of F-specific phages. The concentration of F-specific RNA phages is relatively low (10^1 to 10^3/g) in free-roaming or wild animals (13). The highest numbers are in animal feces from animal husbandry operations (up to a mean of 10^6/g). Hence, feces from wild animals do not appear to be an important source of F-specific RNA phage.

In contrast, domestic sewage is an abundant source of F-specific RNA phages. The concentration of F-specific RNA phages is somewhat lower than that of somatic coliphages by a factor of 2 to 8 and ranges from 400 to 40,000 per milliliter (19; figure 45-1). Similar numbers are found in wastewater from hospitals, pig slaughtering operations, and poultry processing plants (18). The higher numbers of F-specific RNA phages found in wastewater than feces suggest that these phages must be able to multiply in sewage. Multiplication of bacteriophage GA in pasteurized sewage and river and/or seawater has been demonstrated at a temperature of 20°C, but only if the host strain had been pre-grown in broth at 37°C (20). Hence, replication seems to be restricted to environments with direct fecal contamination. As phage and host bacteria are present at concentrations of only about 10,000 per milliliter, it also is not likely that multiplication will take place at environmental sites other than the sewage system. Thus, particularly because of their greater abundance within sewage, the presence of F-specific RNA phages in water is more an index of sewage pollution than just of fecal contamination (23).

Bacteriophages Infecting *Bacteroides fragilis*

Bacteroides fragilis is an anaerobic bacterium that has as its major ecological niche the human intestinal tract, and it has been argued that the same would be true for its phages (29). Phages belonging to the family *Siphoviridae*, with long, flexible tails (double-stranded DNA, long noncontractile tails, and capsids up to 60 nm) are the most common infecting *B. fragilis* (27). Phages infecting this bacterium attach to the cell wall. *B. fragilis* strains differ widely in the numbers of phages that they recover from domestic sewage.

The isolation frequency of *B. fragilis* phages in human feces is relatively low: 0 to 15% of the samples (9). *B. fragilis* strain RYC 2056 has been reported to recover phage from 28% of human stool samples (34). This host has also been shown to isolate phage in 30% of stool samples from pigs. Bacteriophages infecting *B. fragilis* strain HSP40 have been isolated from 10–13% of human stool samples, but not from animal feces (36). Tartera and Jofre (36) indicate that in positive stool samples the concentration of *B. fragilis* may range up to 10^8 per gram. Kai et al. (30) reported a range between 10^2 and 10^5 per gram, and irregular patterns of shedding when a single individual was studied over time. Bacteriophages infecting *B. fragilis* strain RYC2056 are found in domestic sewage (Europe, Africa, and America) in concentrations ranging from 10^2 to 10^3 per milliliter. This is one order of magnitude less than the concentration of F-specific RNA phages. The ratio of F-specific RNA and somatic coliphages is fairly constant in domestic sewage.

Bacteriophages as Indexes of Sewage Pollution

The preceding review suggests that at least three groups of bacteriophages infecting enteric bacteria are consistently found in sewage effluents. Thus, the three groups of phages may be considered as indexes of sewage pollution. On average, all three groups are more abundant in raw sewage than most enteric pathogens (27). No individual group fits the model of an ideal index organism (table 45-5) but they do aid in better defining the quality of water than traditional bacterial indicators.

Bacteriophages as Indexes of Human Enteric Viruses

The three groups of bacteriophages discussed have been proposed as potential indexes of the presence of human enteric viruses (i.e., enteroviruses, hepatitis A virus, Norwalk virus, etc.) on the basis of similar nucleic acid

Table 45-4 Commonly Used Organisms for Assessment of Water Quality

Host	Strain	Strengths	Weaknesses	Stability
S. typhimurium (F$^-$)	WG45	Detects somatic *Salmonella* phages	Shows only somatic attack	
S. typhimurium (F$^+$)	WG49	Reported to be selective for F-specific RNA phages. Low rate of F$^-$ plasmid segregation. Kanamycin and nalidixic acid resistant	Not specific to F-specific RNA phages: also susceptible to attack by *Salmonella* somatic phages and F-specific DNA phages. Somatic *Salmonella* phages cause major interference	An unstable strain that unpredictably loses its ability to plaque F-specific phages
E. coli (F$^-$)	CN, CN13	Nalidixic acid resistant strain		
E. coli (F$^-$)	K-12		Show somatic attack	
E. coli (K-12 F$^+$)	WG21, A/λ, Q13		Susceptible to F-DNA phage attack. Also, plaque somatic T phages. Highly inefficient for enumeration of naturally occurring F-specific RNA phages	
E. coli (F$^-$)	B		Produces plaque counts 5–6 times lower than other phages used for environment assay. Also, plaque somatic T phages	
E. coli (F$^-$)	C	More plaques, highest counts. Nalidixic acid resistant. Most suitable for isolating DNA somatic phages, especially temperate phages		Plaque somatic T phages
E. coli (F$^+$)	C-3000		May be infected by some somatic coliphages. Majority of phages were somatic	
E. coli (K-12 F$^+$)	W3110			
E. coli	R AMP, RR	Ampicillin and streptomycin resistant. Gives the highest % of detection for F-specific RNA phages	Low counts and susceptible to F-specific DNA phage attack	*E. coli* RR, stable

Modified from Leclerc et al. (32).

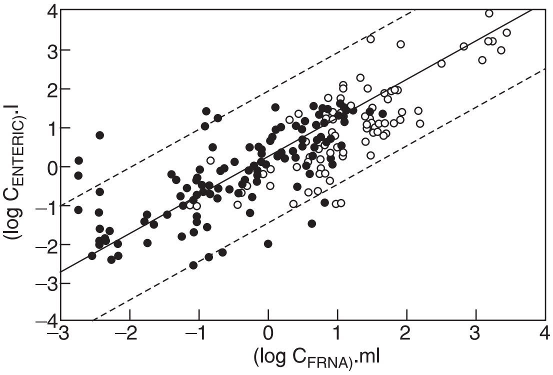


Figure 45-1 Summary of phage densities in fresh waters and sewage. Scatterplot (filled circles), regression line (unbroken line), and 95% confidence intervals (dashed line) of concentrations of enteric viruses in relation to concentrations of F-specific RNA bacteriophages in river and lake water in the Netherlands. For comparision, similar data on raw and treated sewage are also shown (open circles) (19).

composition and morphology. To be useful as index organisms they should have an ecology similar to that of the human enteric viruses. Havelaar (15) listed a number of general criteria for the ideal enteric virus model in water: it should (i) occur exclusively and consistently in human feces, (ii) not occur in animal feces, (iii) not multiply in natural waters, (iv) outnumber human viruses in fecally polluted water by several orders of magnitude, (v) behave like human viruses in fecally polluted water treatment processes, and (vi) be detectable by simple, inexpensive, and rapid methods.

All three groups of bacteriophages considered here occur in human feces. Somatic phages are frequently found in animal feces, but F-specific phages and phages of *B. fragilis* occur infrequently and in lower numbers in the feces of animals. Only the phages of the anaerobe *B. fragilis* seem unlikely to replicate in the environment. The literature indicates that the probability of replication in the environment is greater for somatic coliphages for than F-specific RNA bacteriophages.

Few studies have been conducted on establishing a correlation between human enteric viruses and bacteriophages. The most extensive data on bacteriophages in relation to enteric viruses in water was presented by Havelaar et al. (19). These authors collected data from surveys conducted over a period of 7 years in different types of treated and untreated wastewater, and in fresh waters from the Netherlands. The enteric virus concentrations varied widely between 0.001 and 570 per liter. Bacterial model organisms (fecal coliforms and fecal streptococci) were significantly correlated with enteric viruses in river water and coagulated secondarily treated sewage effluent, but relatively low numbers were found in disinfected effluents and relatively high numbers in surface water open to nonhuman fecal pollution. F-specific RNA bacteriophages were also highly correlated with enteric viruses in all environments studied except for raw (untreated) and biologically (activated) treated sewage. Numerical relationships were consistent over a whole range of different environments; the regression equations of the F-specific RNA bacteriophages on enteric viruses in river water and lake water were statistically equivalent (figure 45-1). On average, a concentration of enteroviruses of 1 per 10 liters would correspond to a concentration of F-specific bacteriophage of 0.2 per milliliter.

Jiang et al. (26) found that somatic coliphages could not be correlated with the presence of human enteric viruses in storm-water runoff impacting coastal waters of Southern California. The presence of human enteric viruses was significantly correlated with F-specific RNA bacteriophages, however. Coliform bacteria, fecal coliforms, and enterococci, on the other hand, did not correlate with the presence of the human adenoviruses.

Recent research also suggests that F-specific RNA bacteriophages are useful indicators of enteric virus contamination of shellfish (oysters, clams, mussels). In a study in England it was found that F-specific RNA bacteriophage concentrations in oysters were strongly associated with harvest-area fecal pollution and with shellfish-associated disease outbreaks (6). Bacteriophage contamination exhibited a marked seasonal trend that

Table 45-5 General Features of Bacteriophage Proposed in Water Quality Assessment

Feature	Somatic coliphages	F-specific RNA phages	Phages of *B. fragilis*
Homogeneity of the group	+	+++	+++
Occurrence and concentration in human feces	++	+	+
Occurrence and concentration in animal feces	+++	+	+
Occurrence and concentration in domestic sewage	+++	+++	+++
Probability of replication in the environment	++	+	−
Resistance to inactivation in the environment	++	+	+++

Modified from Jofre (27).
+++, high; ++, intermediate, + low; −, very low.

was consistent with the trend of oyster-associated gastroenteritis in the United Kingdom.

Stetler (35) monitored the occurrence of somatic coliphages and enteroviruses in river water and at various stages in a conventional drinking water treatment plant. The coliphages could be detected in the source water by direct inoculation, and sufficient coliphages were detected in enterovirus concentrates to permit coliphage levels to be followed through the different water treatment processes. Statistical analysis of the data indicated that enterovirus isolates were better correlated with coliphages than with total coliforms, fecal coliforms, fecal streptococci, or standard plate count bacteria.

Jofre et al. (28) found a correlation between the numbers of *B. fragilis* bacteriophage, enteroviruses, and rotaviruses in sewage-polluted marine sediments. The ratios between the phages and either enteroviruses or rotaviruses in the marine sediments were similar to the ratios found in sewage, suggesting that they have a similar fate in marine sediments. Ganzter et al. (10) studied the relative occurrence and concentration of enteroviruses — by cell culture infectivity and by polymerase chain reaction — and of coliphages in treated secondary sewage effluents. They found a significant correlation between the concentration of somatic coliphages or *B. fragilis* phages and the presence of infectious enteroviruses or the presence of enterovirus genomes.

Bacteriophages as Indexes of Human and Animal Pollution

The identification of the sources of fecal contamination is important in watershed management. For example, pollution of streams may originate from multiple sources, such as birds, septic tanks, or farm animals grazing in the watershed. Knowing the sources allows for more effective control measures. Serotyping of F-specific RNA bacteriophages has allowed them to be divided into four groups. Serotypes II and III have mainly been isolated from human feces, whereas serotypes I and IV are usually found in animal feces (7). It has also recently been shown that the subgroups can be grouped into four main genotypes, which, with few exceptions, show overall comparability with the serotypes. Probes for each genotype allow plaque hybridization and thereby study of the distribution of subgroups isolated from water samples. Subgroups II and III predominate in water samples contaminated with human fecal pollution and subgroups I and IV predominate in animal feces and in water contaminated with animal feces (2, 22). Griffin et al. (13) used geneotyping to determine that animals were the source of fecal contamination in a Florida spring.

Conclusions

Bacteriophages of enteric bacteria have been proposed as indicators of the presence of fecal pollution, sewage, human enteric viruses, and human and/or animal fecal pollution. Each of the three groups of bacteriophages that has been studied has advantages and disadvantages (table 45-5). Somatic coliphages are the most abundant, and the methods for their detection are the most simple. However, they are a heterogeneous group and it has been reported that they may replicate outside the gut, although this has received only limited study. F-specific RNA bacteriophages are second in abundance and are a homogeneous group. Their method of detection is also simple and they are similar in morphology and nucleic acid content to many of the human enteric viruses. However, they do not survive as well as some enteric viruses in natural waters (27). The bacteriophages of *B. fragilis* are present least often and when present occur in lower numbers than the other model bacteriophages. They are a homogeneous group and there is no evidence that they replicate outside the gut. The fact that their method of detection requires anaerobiosis and that they are present in low numbers are drawbacks to their use as indicator organisms.

No one universal indicator of microbial water quality has ever been found. It is clear that in the future multi-indicators of water quality are likely to come into use. The bacteriophages discussed here will likely play a role in water quality assessment and tracking of contaminant sources. The recent development of standardized methods (24, 25) should lead to more research on potential application of bacteriophages as model organisms.

References

1. Ashbolt, N. J., W. O. K. Grabow, and M. Snozzi. 2001. Indicators of microbial water quality, pp. 259–315. *In* L. Fewtrell and J. Bartram. (eds.) Water Quality: Guidelines, Standards and Health. IWA Publishing, London.
2. Chung, H., L. A. Jaykus, G. Lovelace, and M. D. Sobsey. 1998. Bacteriophages and bacteria as indicators of enteric viruses in oysters and their harvest waters. Water Sci. Technol. 38:37–44.
3. Coetzee, O. J. 1962. Bakteriophagen als Indikator fäkaler Wassereinigung. Gesundheits Ingenieur 12:371–372.
4. Dhillon, T. S., Y. S. Chang, S. M. Sun, and W. S. Chau. 1970. Distribution of coliphages in Hong Kong sewage. Appl. Microbiol. 20:1122–1125.
5. Dhillon, T. S., E. K. S. Dhillon, H. C. Chau, W. K. Li, and A. H. C. Tsang. 1976. Studies on bacteriophage distribution: virulent and temperate bacteriophage content of mammalian feces. Appl. Environ. Microbiol. 32:68–74.
6. Dore, W. J., K. Henshilwood, and D. N. Lees. 2000. Evaluation of F-specific RNA bacteriophage as a candidate human

enteric virus indicator for bivalue molluscan shellfish. Appl. Environ. Microbiol. 66:1280–1295.

7. Furuse, K. 1987. Distribution of coliphages in the environment: general considerations, pp. 87–124. *In*: S. M. Goyal, C. P. Gerba, and G. Bitton (eds.) Phage Ecology. Wiley, New York.
8. Furuse, K., A. Ando, S. Oswa, and I. Watanabe. 1981. Distribution of ribonucleic acid coliphages in raw sewage in South and East Asia. Appl. Environ. Microbiol. 41:995–1002.
9. Gantzer, C., J. Henry, and L. Schwartzbrod. 2002. *Bacteroides fragilis* and *Escherichia coli* bacteriophages in human feces. Int. J. Hyg. Environ. Hlth. 205:325–328.
10. Gantzer, C., A. Maul, J. M. Audic, and L. Schwartzbrod. 1998. Detection of infectious enteroviruses, enterovirus genomes, somatic coliphages, and *Bacteroides fragilis* phages in treated wastewater. Appl. Environ. Microbiol. 64:4307–43012.
11. Geldreich, E. E. 1978. Bacterial populations and indicator concepts in feces, sewage, stormwater and soil wastes, pp. 51–98. *In* G. Berg (ed.) Indicators of Viruses in Water and Food. Ann Arbor Science, Ann Arbor, Mich.
12. Gerba, C. P. 1987. Phage as indicators of fecal pollution, pp. 197–210. *In* S. M. Goyal, C. P. Gerba, and G. Bitton. Phage Ecology. Wiley, New York.
13. Griffin, D. W., R. Stokes, J. B. Rose, and J. H. Paul. 2000. Bacterial indicator occurrence and the use of an F(+) specific RNA coliphage assay to identify fecal sources in Homosassa Springs, Florida. Microb. Ecol. 39:56–64.
14. Havelaar, A. H. 1993. A Bacteriophage Standard for Bathing Beaches. Final Report. National Institute of Public Health and Environmental Protection, Bilthoven, The Netherlands.
15. Havelaar, A. H. 1993. Bacteriophages as models of human enteric viruses in the environment. ASM News 59:614–619.
16. Havelaar, A. H., K. Furuse, and W. M. Hogeboom. 1986. Bacteriophages and indicator bacteria in human and animal feces. J. Appl. Bacteriol. 60:255–262.
17. Havelaar, A. H., and W. M. Hogeboom. 1984. A method for the enumeration of coliphages in sewage and sewage-polluted waters. Antonie van Leeuwenhoek 49:387–397.
18. Havelaar, A. H., W. M. Hogeboom, and R. Pot. 1984. F-specific RNA bacteriophages in sewage: methodology and occurrence. Water Sci. Technol. 17:645–655.
19. Havelaar, A. H., M. Olphen, Y. C. Olphen, and Y. C. van Dorst. 1993. F-specific RNA-bacteriophages are adequate model organisms for enteric viruses in fresh water. Appl. Environ. Microbiol. 59:2956–2962.
20. Havelaar, A. H., and W. M. Pot-Hogeboom. 1988. F-specific RNA-bacteriophages as model viruses in water hygiene: ecological aspects. Water Sci. Technol. 20:399–407.
21. Havelaar, A. H., W. M. Pot-Hogeboom, K. Furuse, R. Pot, and M. P. Hormann. 1990. F-specific RNA bacteriophages and sensitive host strains in feces and wastewater of human and animal origin. J. Appl. Bacteriol. 69:30–37.
22. Hsu, F. C., Y. S. Shieh, J. van Duin, M. J. Beekwilder, and M. D. Sobsey. 1995. Geneotyping male-specific RNA coliphages by hybridization with oligonucleotide probes. Appl. Environ. Microbiol. 61:3960–3966.
23. IAWPRC Study Group on Health Related Water Microbiology. 1991. Bacteriophages as model viruses in water quality control. Water Res. 25:529–545.
24. ISO. 1995. Water Quality. Detection and Enumeration of Bacteriophages. Part 1: Enumeration of F-specific RNA Bacteriophages. ISO/DIS10705-1. International Organization for Standardization, Geneva, Switzerland.
25. ISO. 1999. Water Quality. Detection and Enumeration of Bacteriophages. Part 2. Enumeration of Bacteriophages Infecting *Bacteroides fragilis.* ISO/DIS10705-4. International Organization for Standardization, Geneva, Switzerland.
26. Jiang, S., R. Noble and W. Chu. 2001. Human adenoviruses and coliphages in urban runoff-impacted coastal waters of Southern California. Appl. Environ. Microbiol. 67:179–184.
27. Jofre, J. 2002. Bacteriophage indicators, pp.354–363. *In* G. Bitton (ed.) Encyclopedia of Environmental Microbiology. Wiley, New York.
28. Jofre, J., M. Blasi, A. Bosch, and F. Lucena. 1989. Occurrence of bacteriophages infecting *Bacteroides fragilis* and other viruses in polluted marine sediments. Water Sci. Technol. 21:15–19.
29. Jofre, J., A. Bosch, R. Lucena, R. Girones, and C. Tartera. 1986. Evaluation of *Bacteroides fragilis* bacteriophages as indicators of the virological quality of water. Water Sci. Technol. 18:167–173.
30. Kai, S., S. Wantanabe, K. Furuse and A. Furuse, and A. Osawa. 1985. *Bacteroides* bacteriophages isolated from human feces. Microbiol. Immunol. 29:895–899.
31. Kott, Y. 1966. Estimation of low numbers of *Escherichia coli* bacteriophage by use of the most probable number method. Appl. Microbiol. 14:141–144.
32. Leclerc, H., S. Edberg, V. Pierzo, and J. M. Delattre. 2000. Bacteriophages as indicators of enteric viruses and public health risk in groundwaters. J. Appl. Microbiol. 88:5–21.
33. Osawa, S., K. Furuse and I. Watanabe. 1981. Distribution of ribonucleic acid coliphages in animals. Appl. Environ. Microbiol. 41:164–168.
34. Puig, A., N. Queralt, J. Jofre and R. Araujo. 1999. Diversity of *Bacteroides fragilis* strains in their capacity to recover phages from human and animal wastes and from fecally polluted wastewater. Appl. Environ. Microbiol. 65:1772–1776.
35. Stetler, R. E. 1984. Coliphages as indicators of enteroviruses. Appl. Environ. Microbiol. 48:668–670.
36. Tartera, C., and J. Jofre. 1987. Bacteriophages active against *Bacteroides fragilis* in sewage–polluted waters. Appl. Environ. Microbiol. 24:1632–1637.
37. Woody, M. A., and D. O. Cliver. 1997. Replication of coliphage Q beta as affected by host cell number, nutrition, competition from insusceptible cells and non-F-RNA coliphages. J. Appl. Microbiol. 82:431–440.

46

The Use of Phage as Diagnostic Systems

CATH REES

The most common use of bacteriophage in detection methodology is phage typing. When developing panels of phage which will discriminate between isolates of the same species on the basis of lytic spectrum, phage are chosen specifically because of their narrow (and therefore discriminatory) host range. Since this is the use of phage most commonly encountered, a general impression has been created that phage have a narrow host range. Hence, one of the most commonly voiced criticisms of phage-based detection methods is that the host range of phage — and therefore any test — is limited, but this is far from the truth. When phage are selected for detection of species or of whole genera, those with the widest host range are chosen. A classic example of this is the phage Felix 01, first described as a broad-host-range phage infecting *Salmonella enterica* by Cherry et al. (8). Subsequent extensive studies of its host range have shown that it will usually infect more than 95% of *Salmonella* isolates (see 16 for a review). Similarly, in the Gram-positive genera, listeriaphage A511 was reported to infect 95% of the two major serotypes of *Listeria* associated with human disease (serotypes 1/2a and 4b; 17). More recently, systematic searches for broad-host-range lytic phage have demonstrated that they can be readily isolated from natural communities where the different bacterial genera are likely to be found (14). Hence, phage-based detection tests have been successfully developed, although few have met with commercial success.

One reason for this lack of commercial success is that the tests are rarely 100% effective —there will always be mutants in a population that are phage-resistant. In the case of reporter phage assays, however (see below), it has been shown that host cells that will not propagate a phage will still support infection and expression of the reporter gene (i.e., the infection host range is wider than the replication host range; see 16), and this fact increases the versatility of reporter phage tests. It must also be remembered that even using classical microbiological identification techniques, atypical isolates are often reported which would be recorded as false negatives, and in DNA-based detection methods there is often a concern about detecting both dead cells and cells containing cryptic genes. Therefore, the criticisms leveled at the phage-based assays do seem to be unfairly biased by this general belief that "phage have a narrow host range" and this has hampered their development. In practice, phage-based assays have had to offer some additional benefit before being adopted in favor of other rapid molecular-biological detection tests. Those that have been successfully developed are described below, along with some newer concepts that are still at the developmental stage that may represent future applications of phage in diagnostic systems.

Reporter Phage

Of the several different types of phage-based detection tests developed to date, the most widely described thus far has been recombinant reporter phage. The idea that phage could be used to introduce reporter genes into a bacterial cell had been well established by the development of phage Mud-*lac* as a genetic tool. However, it was Ulitzer and Kuhn (35) who first proposed the idea of using the expression of the reporter genes following infection by phage specifically as a rapid method of bacterial detection (figure 46-1). The effectiveness of this methodology was first demonstrated by simply using phage-based cloning vectors engineered to contain the complete bacterial bioluminescence (*lux*) operon, and by using such constructs it was shown that as few as 10 *Escherichia coli* cells could be detected (34). In addition this group carried out random mutagenesis of wild-type phage genomes using Tn*10* transposons carrying the bacterial luciferase genes (*luxAB*) and recovered recombinant phage from bioluminescent plaques. These simple constructs have been shown to successfully detect the presence of enteric pathogens in food samples (7, 15), but the most successful development of reporter phage has been for *Mycobacterium tuberculosis*.

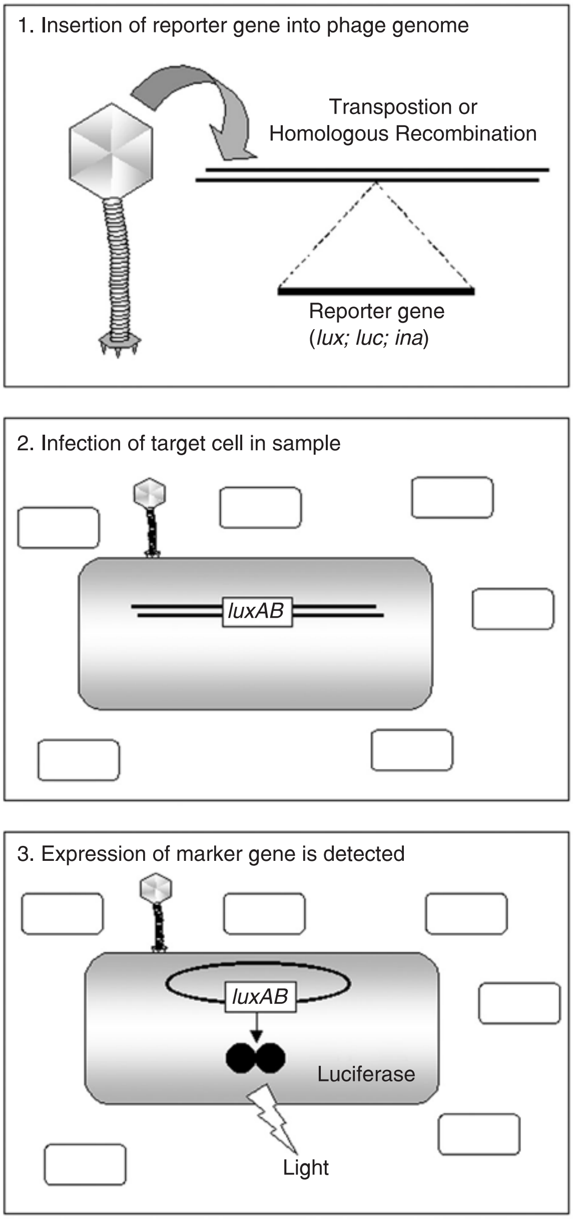

1. Insertion of reporter gene into phage genome
Transpostion or
Homologous Recombination
Reporter gene
(lux; luc; ina)
2. Infection of target cell in sample
luxAB
3. Expression of marker gene is detected
luxAB
Luciferase
Light

Figure 46-1 Reporter phage assay. 1: Reporter genes (e.g., bacterial luciferase, *lux*; firefly luciferase, *luc*; ice nucleation protein, *ina*) are introduced into phage genomes either by transposition or via homologous recombination between wildtype phage and cloned phage sequences carrying the reporter gene. Recombinant phage are purified from the primary phage lysate by plaque purification and phenotypic screening for the expression of the reporter gene. 2: Purified Reporter phage (e.g., *lux* phage) are used to infect target cells (drawn as large, shaded cell) in a mixed sample (other bacterial present represented by unshaded cells). Phage DNA is injected into the host cell and replication begins. 3: Phage genes are expressed along with the inserted reporter gene. The signal from the reporter gene (e.g., light) is detected indicating the presence of a cell which is sensitive to phage infection.

The first reporter phage to be developed for the detection of Mycobacteria was based on the lytic phage, TM4, in this case using firefly luciferase (*luc*) as the reporter gene (13). However, the rapidity of the lytic cycle meant that only limited amounts of luciferase protein were produced and the limit of detection using this construct was found to be 10^4 mycobacterial cells (13). A lytic phage had been chosen because when predicting the "ideal" features of reporter phage it was reasoned that a lytic phage would be the most effective, as all infecting phage would be actively replicating leading to high-level expression of the reporter genes. In fact, this did not prove to be a crucial factor in the successful development of reporter phage for Mycobacteria.

When the same group constructed a reporter phage based on the temperate phage, L5, they found that a constitutive promoter was fortuitously generated when the phage integrated into the chromosome, resulting in prolonged expression and accumulation of the luciferase protein and thus enhanced amplification of the signal generated. This allowed the limit of detection to be reduced to approximately 10^2cells after a 40 hour incubation period, or 10^3 cells after 20 hours (29). However, the limited host range of the L5 phage meant that it was not directly applicable for the detection of *M. tuberculosis* in a commercial test. Further work has been carried out to improve the TM4-based phage, including changing the site of insertion of the *luc* gene in the phage genome and isolating various spontaneous mutants of the parent phage. As a result, derivatives of TM4 can now detect as few as 120 *Mycobacterium bovis* BCG after only a 12 hour incubation period (6), providing a dramatic improvement in detection times over standard culture methods for this slow-growing organism. These phage have been used in combination with further refinements of the assay conditions and shown to successfully detect Mycobacteria in smear-positive sputum samples within 24–48 hours (26).

In each of the cases described above, a random approach was taken to construction of the reporter phage, with the formation of plaques by progeny phage produced following the recombination/transposition event being used to identify viable constructs. This approach has also facilitated studies of the phage genomes, leading to the identification of nonessential regions (see 25, 29). However, a more structured approach has also been employed by first characterizing regions of phage genome and then integrating the reporter gene into a specific locus. This approach was used by Loessner et al. (20) who introduced the *luxAB* genes into a late region of a virulent listeriaphage A511 (see chapter 37 for review of *Listeria* phage as well as a brief discussion of this reporter phage work). A511 is a member of the *Myoviridae* family and has a double-stranded DNA genome of approximately 116 kbp (18). Studies of the capsid proteins had allowed the late regions containing the genes for the major capsid and tail sheath proteins to be identified (19). The *luxAB* genes were introduced into a cloned fragment of phage DNA to create operon fusions with the phage structural genes without disrupting any of the original gene structures. Recombinant phage were recovered from lysates following infection of *Listeria* cells carrying the plasmid constructs with wild-type phage. These progeny phage were plated on a lawn of propagating bacteria and plaques screened for a bioluminescent phenotype when the lawn was exposed to aldehyde vapor. Double-recombinant phage, which contained only the additional *luxAB* genes and no vector sequences, were found at an unexpectedly high frequency (5×10^{-4}; 20).

This reporter phage was evaluated for its ability to successfully detect the presence of *Listeria monocytogenes* in different types of food samples (21). From this study it was clear that, as for many rapid methods applied to food systems, the nature of the food matrix and the competing microflora is critical to the sensitivity of the test. Generally 1–10 cells per 25 g food sample was detectable in relatively simple foods matrices, but the threshold of detection was nearer 200 cells per 25 g in more complex samples. This study highlighted one of the main advantages of a phage-based test: the specificity of the host–phage interaction allows the detection of low numbers of specific bacteria in a mixed population without the need for purification to homogeneity by successive rounds of enrichment and selective plating.

The strategy used to construct these *Listeria lux* phages requires a minimum amount of genome analysis to allow successful insertion of the reporter genes while retaining all phage essential functions. The strategy is generically applicable to all phage that can accommodate the insertion of the marker gene without exceeding the packaging constraints of the phage. In this case the *luxAB* reporter genes are 2.1 kbp in size. Similarly, the firefly *luc* gene is 2.3 kbp in size and the bacterial ice nucleation protein, which has also been used to construct reporter phage (39), is even larger at 3.4 kbp. The directed approach to phage construction is only effective if the packaging constraints of the phage are sufficient to allow incorporation of the marker genes into the phage genome. If the marker gene is too large to be accommodated within the phage genome, then compensating deletion of nonessential genes is necessary. However, such deletion requires further analysis of the phage genome so that appropriate redundant regions can be identified.

The choice of reporter gene relates to the sensitivity with which the signal generated is to be detected. For instance, it is reported that only one molecule of ice nucleation protein needs to be synthesized to allow a positive test result (BIND assay: bacterial ice nucleation detection; 38). Hence this was used in a commercial test developed for the rapid detection of *Salmonella* as even low level expression is readily detected. The sensitivity of this assay comes from a 2-fold amplification of signal. Firstly, the reporter gene (ice nucleation protein) is expressed at high levels in the infected cell and these proteins become localized in the outer membrane. Secondly, the presence of the protein is detected using a phase-sensitive fluorescent dye which changes color as the buffer freezes. Hence, when samples are cooled, only buffer solutions containing transfected salmonellae freeze, causing the dye to change color. This assay allowed the detection of samples containing only 2 cells/ml *Salmonella enteritidis* in a buffer system within 3 hours. The sensitivity of the assay was further increased by using it in combination with salmonellae-specific immunomagnetic bead separation (12).

All the reporter phage described so far have been viable phage with a full complement of genes to allow phage replication. In light of public concerns about the release of recombinant microorganisms, use of these types of phage will be necessarily be restricted to specialized laboratories. More recently, Kuhn et al. (16) have described the construction of "locked" phage that do not produce productive phage particles unless grown on a specially engineered host strain. First they isolated double amber mutants of the *Salmonella* phage Felix 01. These would not grow on *Salmonella enterica* LT2 and would only propagate on the sup^{+} *Salmonella* strain K772 which can suppress such amber mutations. Being double amber phage mutants, reversion rates were very low (10^{-8}–10^{-9} per generation). Genetic analysis of the phage genome was carried out to identify a region of approximately 3 kbp which contained both essential and a nonessential genes. This region was replaced with a DNA fragment containing the bacterial luciferase genes (*luxAB)* and the *supF* gene by homologous double recombination. For the propagation of such phage, a sup^{-} host was used which contained the missing segment of the Felix −01 genome on a plasmid construct to provide the essential phage gene functions *in trans*. In this case, recombinant phage were only rarely recovered, which was thought to be due either to a limitation in the recombination event (only short homologous flanking regions were present in the plasmid construct) or possibly to a feature of the phage biology if degradation of host (including plasmid) DNA occurred following phage infection.

The phage produced in this way are genetically "locked," as they need to maintain the *luxAB-supF* segment to allow them to propagate in wild-type *Salmonella*, and, due to the missing essential gene, will only successfully replicate in a strain containing the complementing plasmid. Following infection of other *Salmonella* strains, the luciferase genes are successfully expressed, allowing detection of the infected cell, but no viable phage particles are produced at the end of the infection cycle. Although this method requires many more manipulations and far more extensive genetic analysis of individual phage, the fact that the reporter phage produced are effectively nonviable when used in the assay should go some way to address the concerns of environmental release of genetically modified organisms.

Phage Amplification Assay

The major draw back of the reporter phage technology is the fact that genetic engineering is required for each new phage to be employed. The difficulty of generating such phage means that new phage cannot quickly be developed for the detection of either new genera or new species/subspecies. The emergence of new virulent subspecies (such as the appearance of *Escherichia coli* O157:H7) or changes in the requirement to detect human pathogens (such as the identification of *Campylobacter* as a significant human pathogen) means that rapid detection methods need to be easily adapted to detect these new groups. A method that fulfils this requirement is phage amplification technology. This test uses only wild-type phage and the endpoint of detection is the formation of a plaque. No gene engineering is required. Therefore there are no containment concerns and the technology uses standard microbiological methods. Hence, staff carrying out routine testing do not need high levels of training in molecular microbiological methods and the methodology can easily be established as part of more traditional microbiological testing regimes.

The assay was first described by Stewart et al. (31, 33) who named it termed phage amplification. A similar methodology has been described by Wilson et al. (22, 36) but termed phage amplified biologically assay (PhaB). The word "amplification" is used to describe these assays because the primary phage infection event which signals the presence of the target bacterial cell is only detected when the progeny phage are allowed to replicate and develop into a plaque (figure 46-2). The test begins by adding phage specific for the target bacterium to the test sample. Time is allowed for the phage to bind to the host cell and to enter the eclipse phase with the nucleic acid delivered into the host cell cytoplasm. At this point a virucide is added to destroy all those phage which have not successfully infected a bacterial cell. After a short period of incubation to allow complete killing of all the free phage particles, the virucide is neutralized and the sample mixed with a laboratory strain (termed the "helper" bacteria) which is known to support the replication of the phage used. This mixture is then spread over the surface of an agar plate as a soft agar overlay. Once the replication of the eclipsed phage is complete, cell lysis of the target

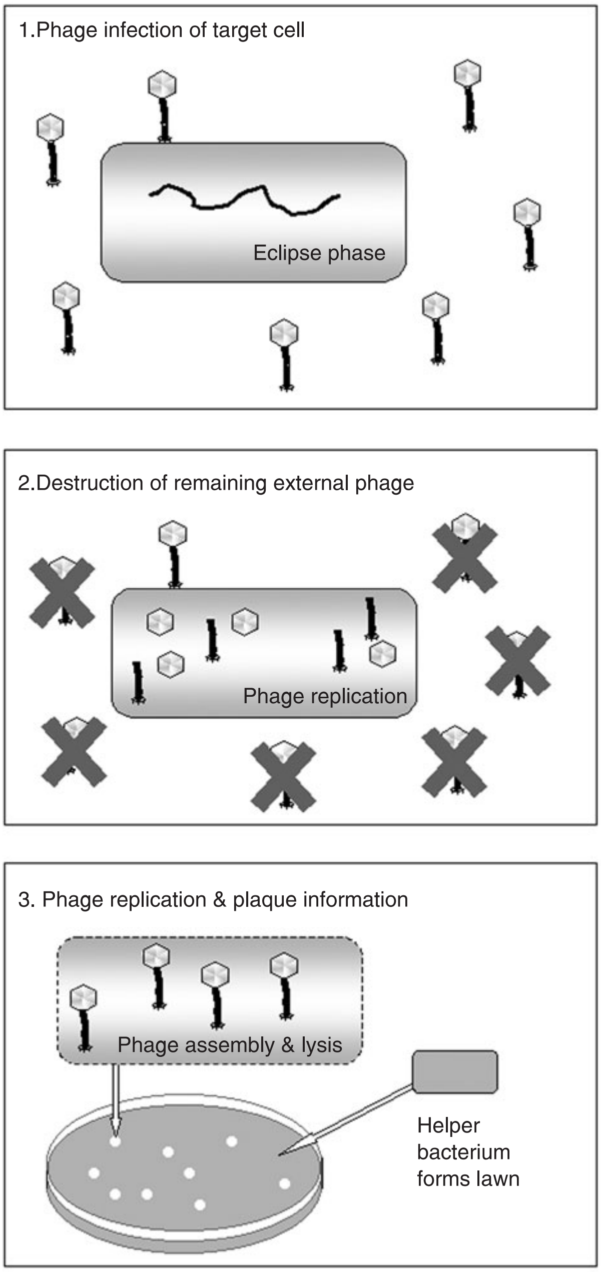

1.Phage infection of target cell
Eclipse phase
2.Destruction of remaining external phage
Phage replication
3. Phage replication & plaque information
Phage assembly & lysis
Helper bacterium forms lawn

Figure 46-2 Phage amplification (or PhaB) assay. 1: Phage are used to infect target cells in sample and time is allowed for infection to proceed to the eclipse phase. 2: A virucide is added that will inactivate phage which have not infected but does not affect the viability of the host cell. Phage replication continues in the infected cells. 3: The virucide is neutralised and the infected cells are mixed with a large excess of helper cells, known to support the replication of the phage. The whole sample is mixed with soft agar to form a bacterial lawn. Replication of the infected cell is completed and cell lysis occurs. Released phage particles infect helper cells in the lawn and plaques are formed. Each plaque represents one infected cell in the original sample.

bacterium occurs and the progeny phage are released. These can then go on to infect the helper bacteria present in the lawn and a plaque develops at each site where originally there was an infected bacterial cell in the sample tested. Hence, each plaque represents the presence of one target bacterium in the original sample.

This technology has been shown to be effective for the detection of a diverse range of bacterial pathogens such as *Listeria, Campylobacter, Pseudomonas, Salmonella,* and *Escherichia coli*, but to date it has been commercially developed only for the detection of *Mycobacterium tuberculosis* (FAST*Plaque*TB^{TM}; see 24). When testing for the presence of slow-growing organisms, such as *M. tuberculosis*, the great advantage of these assays is that the target organism and the helper organism do not need to be the same, as long as the phage can infect both cell types. Thus, in the *M. tuberculosis* assay, the helper organism chosen is *M. smegmatis*, which can form lawns within 12–48 hours rather than the 8 weeks required for *M. tuberculosis* cells to grow into visible colonies. For fast-growing bacterial pathogens there is no obvious advantage to the technique, as colonies can develop using standard microbiological detection methods within 12 hours. However, this does not take account of the time required to carry out the confirmatory tests which are often needed following the presumptive identification of a bacterial colony. When using the phage amplification procedure, much of this specificity can be built in by choosing phage with a suitable host range, thus achieving an overall reduction in detection time.

Antibiotic Sensitivity Testing

Both the reporter phage technology and the phage amplification technique have been used to develop antibiotic sensitivity tests for the pathogens detected by these assays. The basis of this type of test is slightly different depending on which type of phage-based assay is used, but the end point assay is much the same. Simply: if the antibiotic is added and the target cell is sensitive to it, then the phage infection event is not detected as no signal is generated. In the case of the *luc* reporter phage test, no expression of the reporter gene is seen since the cells need to be alive and to carry out both protein and ATP synthesis necessary to produce bioluminescence from the firefly luciferase (6, 13, 27). When using the phage amplification test, simply no plaques would be formed as primary replication of the phage in the target cell is inhibited (1, 10). Hence by comparing the results of the assay with and without the addition of the antibiotic, the resistance (or sensitivity) of a bacterial pathogen can be rapidly determined without the need to wait for the growth of that organism following isolation. As before, the greatest advantage here is for slow growing or fastidious organisms that are difficult to culture, and accordingly tests have been developed and evaluated for determining the antibiotic resistance profile of *M. tuberculosis* strains isolated from patients prior to antibiotic therapy (2, 3, 23).

Phage-Mediated Release of ATP

One of the most commonly used rapid hygiene tests used is the bioluminescent determination of released bacterial ATP (see 30). These assays are based on the fact that there is a linear relationship between the number of photons produced by firefly luciferase and the number of ATP molecules hydrolyzed and the fact that the amount of ATP per bacterial cell in a given growth condition is quite constant (approximately 10^{-15} g per cell). Many commercial companies produce kits which provide the luciferin and luciferase substrates to allow simple measurement of ATP levels using a luminometer, and also included are the lysing reagents required to break open the bacterial cells and release the intracellular ATP for measurement. This, then, is the limitation of the technique in that it gives only a measure of microbial load and no indication of the composition of the microflora.

The ability of phage to lyse only specific cells within a mixed population has been used to add specificity to this test. This was first demonstrated for the detection of *Listeria* using both intact phage (28) and purified phage lysin (32). However, although both groups could add specificity to the test, the problem exists that the practical limit of detection for these assays is approximately 10^4 bacterial cells; below this number the amount of light generated is too low to be detected against background. Obviously this does not give the level of sensitivity required when testing for pathogens, when tests must often demonstrate the absence of that organism from 25 g of sample. The sensitivity of the assay has been improved by amplification of the ATP signal through the measurement of released adenylate kinase (see 9) which amplifies the bioluminescence signal due to its ability to continue generating ATP from ADP in a linear fashion. Using this technique, it has been shown that the limit of detection of phage-based assays can be reduced to less than 10^3 *E. coli* and *Salmonella* cells (5, 40), which compares favorably with other rapid methods of bacterial enumeration such as immunoassays.

Future Horizons

One of the most interesting new ideas put forward for the use of bacteriophage in detection assays is the dual phage technology (37; figure 46-3). In this case it is not the specificity of the host–phage relationship that is exploited. Rather, this technology uses phage to report on the successful binding of an antibody to a specific antigen. The

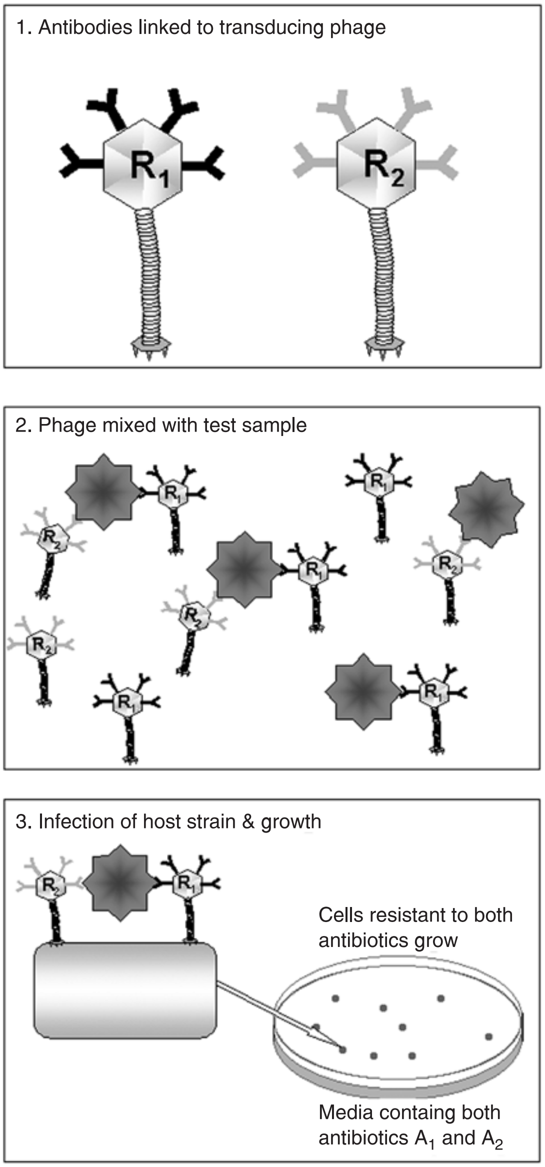

1. Antibodies linked to transducing phage
R1
R2
2. Phage mixed with test sample
3. Infection of host strain & growth
Cells resistant to both antibiotics grow
Media containg both antibiotics A1 and A2

Figure 46-3 Dual phage detection assay. 1: Two transducing phage are used carrying resistance genes (R_1 and R_2) to two different antibiotics (A_1 and A_2). Different antibodies specific for the same antigen are bound to the phage capsids. 2: Phage–antibody complexes are mixed with the test sample and antibody binding allowed to occur. 3: Phage are mixed with a propagating strain at a low multiplicity of infection and plated on agar containing both antibiotics (A_1 and A_2). Only cells which have received both resistance genes grow into colonies, indicating that the phage were physically linked together through antigen binding.

phage used are transducing phage that can confer antibiotic resistance on a host cell following infection. The antibodies are linked to the surface of two different phage particles, each one conferring resistance to a different antibiotic. Antibody-linked phage are mixed with the sample and if the antigen is present then the two different phage types are physically linked closely together. Host cells for the phage are then added to the mixture and the cells plated out on media containing both antibiotics. A low multiplicity of infection of phage is used, so normally the probability of coinfection by the two types of phage is very low. If large numbers of double-resistant colonies appear on the plates, then this indicates that the different types of transducing phage had been located close together while linked to the same antigenic particle via the antibody moieties and this dual infection signals a positive detection event.

In this case there is greater than a million-fold amplification of the signal, as a single infection event grows into a visible antibiotic-resistant colony containing more than 10^8 cells, and thus the sensitivity of the test is great. The ability to use two different antibodies in one detection assay also ensures a high degree of specificity. This dual phage assay has the advantage that no phage engineering is required and it can be applied to any type of antigen for which an antibody exists, not just for the detection of bacterial cells. Trials have been reported showing that the assay could sensitively detect human immunodeficiency virus particles in blood samples (S. Wilson, Microsens Biophage , UK, unpublished data) and it is also intended to develop sensitive dual phage assays for prions.

This type of ingenious application of our understanding of phage biology suggests that many more applications will appear in the future, and the more we understand the biology and genetics of phage, the more applications will emerge. For instance, the fact that phage have evolved to specifically recognize and bind to structures on the bacterial cell surface means that as our knowledge of structural biology increases we can expect to see further applications of this feature. A simple example of how this might be exploited is a report of the use of *Salmonella* phage Felix 01 (also referred to as phage "Sapphire") as a biosorbant to selectively separate cells from suspensions containing other related bacterial cell types (4). We can speculate that once the equivalent of the antibody minimal-recognition unit can be defined (see 11), phage receptors could be used to generate recombinant proteins, or protein complexes, with the desired binding properties — possibly even contributing to the development of novel families of affinity-binding proteins.

References

1. Albert, H., A. Heydenrych, R. Mole, A. Trollip, and L. Blumberg. 2001. Evaluation of FASTPlaqueTB-RIF^{TM}, a rapid, manual test for the determination of rifampicin resistance from *Mycobacterium tuberculosis* cultures. Int. J. Tubercul. Lung Dis. 5:906–911.
2. Albert, H., A. Heydenrych, R. Brookes, R. J. Mole, B. Harley, E. Subotsky, R. Henry, and V. Azevedo. 2002. Performance of

a rapid phage-based test, FASTPlaqueTB$^{\text{TM}}$, to diagnose pulmonary tuberculosis from sputum specimens in South Africa. Int. J. Tubercul. Lung Dis. 6:529–537.

3. Banaiee, N., M. Bodadilla-del-Valle, S. Bardarov, P. F. Riska, P. M. Small, A. Ponce-De-Leon, W. R. Jacobs, G. F. Hatfull, and J. Sifuentes-Osornio. 2001. Luciferase reporter mycobacteriophages for detection, identification, and antibiotic susceptibility testing of *Mycobacterium tuberculosis* in Mexico. J. Clin. Microbiol. 39:3883–3888.
4. Bennett, A. R., F. G. D. Davids, S. Vlahodimou, J. G. Banks, and R. P. Betts. 1997. The use of bacteriophage-based systems for the separation and concentration of *Salmonella*. J. Appl. Microbiol. 83:259–265.
5. Blasco, R., M. J. Murphy, M. F. Sanders, and D. J. Squirrell. 1998. Specific assays for bacteria using phage-mediated release of adenylate kinase. J. Appl. Microbiol. 84: 661–666.
6. Carriere, C., P. F. Riska, O. Zimhony, J. Kriakov, S. Bardarov, J. Burns, J. Chan, and W. R. Jacobs. 1997. Conditionally replicating luciferase reporter phages: improved sensitivity for rapid detection and assessment of drug susceptibility of *Mycobacterium tuberculosis*. J. Clin. Microbiol. 35:3232–3239.
7. Chen, J., and M. W. Griffiths. 1996. *Salmonella* detection in eggs using *lux*$^{+}$ bacteriophages. J. Food Protect. 59:908–914.
8. Cherry, W. B., B. R. Davis, P. R. Edwards, and R. B. Hogan. 1954. A simple procedure for the identification of the genus *Salmonella* by means of a specific bacteriophage. J. Lab. Clin. Med. 44: 51–55.
9. Corbitt, A. J., N. Bennion, and S. J. Forsythe. 2000. Adenylate kinase amplification of ATP bioluminescence for hygiene monitoring in the food and beverage industry. Letts. Appl. Microbiol. 6:443–447.
10. Eltringham, I. J., S. M. Wilson, and F. A. Drobniewski. 1999. Evaluation of a bacteriophage-based assay (phage amplified biologically assay) as a rapid screen for resistance to isoniazid, ethambutol, streptomycin, pyrazinamide, and ciprofloxacin among clinical isolates of *Mycobacterium tuberculosis*. J. Clin. Microbiol. 37:3528–3532.
11. Harris, B. 1999. Exploiting antibody-based technologies to manage environmental pollution. Trends Biotechnol. 17:290–296.
12. Irwin, P., A. Gehring, S. I. Tu, J. Brewster, J. Fanelli, and E. Ehrenfeld. 2000. Minimum detectable level of Salmonellae using a binomial-based bacterial ice nucleation detection assay (BIND). J. AOAC Int. 83: 1087–1095.
13. Jacobs, W. R., R. G. Barletta, R. Udani, J. Chan, G. Kalkut, G. Sosne, T. Kieser, G. J. Sarkis, G. F. Hatfull, and B. R. Bloom. 1993. Rapid assessment of drug susceptibilities of *Mycobacterium tuberculosis* by means of luciferase reporter phages. Science 260:819–822.
14. Jensen, E. C., H. S. Schrader, B. Rieland, T. L. Thompson, K. W. Lee, K. W. Nickerson, and T. A. Kokjohn. 1998. Prevalence of broad-host-range lytic bacteriophages of *Sphaerotilus natans*, *Escherichia coli*, and *Pseudomonas aeruginosa*. Appl. Environ. Microbiol. 64:575–580.
15. Kodikara, C. P., H. H. Crew, and G. S. A. B. Stewart. 1991. Near on-line detection of enteric bacteria using *lux* recombinant bacteriophage. FEMS Microbiol. Letts. 83:261–266.
16. Kuhn, J., M. Suissa, J. Wyse, I. Cohen, I. Weiser, S. Reznick, S. Lubinsky-Mink, G. Stewart, and S. Ulitzur. 2002. Detection of bacteria using foreign DNA: the development of a bacteriophage reagent for *Salmonella*. Int. J. Food Microbiol. 74:229–238.
17. Loessner, M. J., and M. Busse. 1990. Bacteriophage typing of *Listeria* species. Appl. Environ. Microbiol. 56:1912–1918.
18. Loessner, M. J., I. B. Krause, T. Henle, and S. Scherer. 1994. Structural proteins nnd DNA characteristics of 14 *Listeria* typing bacteriophages. J. Gen. Virol. 75:701–710.
19. Loessner, M. J., and S. Scherer. 1995. Organization and transcriptional analysis of the *Listeria* phage A511 late gene region comprising the major capsid and tail sheath protein genes *cps* and *tsh*. J. Bacteriol. 177:6601–6609.
20. Loessner, M. J., C. E. D. Rees, G. S. A . B. Stewart, and S. Scherer. 1996. Construction of luciferase reporter bacteriophage A511::*luxAB* for rapid and sensitive detection of viable *Listeria* cells. Appl. Environ. Microbiol. 62:1133–1140.
21. Loessner, M. J., M. Rudolf, and S. Scherer. 1997. Evaluation of luciferase reporter bacteriophage A511::*luxAB* for detection of *Listeria monocytogenes* in contaminated foods. Appl. Environ. Microbiol. 63:2961–2965.
22. McNerney, R., S. M. Wilson, A. M. Sidhu, V. S. Harley, Z. Al Suwaidi, P. M. Nye, T. Parish, and N. G. Stoker. 1998. Inactivation of mycobacteriophage D29 using ferrous ammonium sulphate as a tool for the detection of viable *Mycobacterium smegmatis* and *M. tuberculosis*. Res. Microbiol. 149:487–495.
23. McNerney, R., P. Kiepiela, K. S. Bishop, P. M. Nye, and N. G. Stoker. 2000. Rapid screening of *Mycobacterium tuberculosis* for susceptibility to rifampicin and streptomycin. Int. J. Tubercul. Lung Dis. 4:69–75.
24. Mole, R. J., and T. W. O'C. Maskell. 2001. Phage as a diagnostic — the use of phage in TB diagnosis. J. Chem. Technol. Biotechnol. 76:683–688.
25. Pearson, R. E., S. Jurgensen, G. J. Sarkis, G. F. Hatfull, and W. R. Jacobs. 1996. Construction of D29 shuttle phasmids and luciferase reporter phages for detection of Mycobacteria. Gene 183:129–136.
26. Riska, P. F., W. R. Jacobs, B. R. Bloom, J. McKitrick, and J. Chan. 1997. Specific identification of *Mycobacterium tuberculosis* with the luciferase reporter mycobacteriophage: use of *p*-*n*itro-α-Actylamino-β-hydroxy Propiophenone. J. Clin. Microbiol. 35:3225–3231.
27. Riska, P. F., Y. Su, S. Bardarov, L. Freundlich, G. Sarkis, G. Hatfull, C. Carriere, V. Kumar, J. Chan, and W. R. Jacobs. 1999. Rapid film-based determination of antibiotic susceptibilities of *Mycobacterium tuberculosis* strains by using a luciferase reporter phage and the Bronx box. J. Clin. Microbiol. 37:1144–1149.
28. Sanders, M. F. 1995. A rapid bioluminescent technique for the detection and identification of *Listeria monocytogenes* in the presence of *Listeria innocua*, pp. 454–457. *In* A. K. Campbell, L. J. Kricka, and P. E. Stanley(eds.) Bioluminescence and Chemiluminescence: Fundamental and Applied Aspects.Wiley, Chichester.
29. Sarkis, G. J., W. R. Jacobs, and G. F. Hatfull. 1995. L5 Luciferase reporter mycobacteriophages — a sensitive tool for

the detection and assay of live Mycobacteria. Mol. Microbiol. 15:1055–1067.

30. Stanley, P. E. 1989. A review of bioluminescence ATP techniques in rapid microbiology. J. Biolumin. Chemilumin. 4:375–380.
31. Stewart, G. S. A. B., S. A. A. Jassim, S. P. Denyer, S. Park, K. Rostas-Mulligan, and C. E. D. Rees. 1992. Methods for rapid microbial detection. PCT Patent WO 92/02633.
32. Stewart, G. S. A. B., M. J. Loessner, and S. Scherer. 1996. The bacterial *lux* gene bioluminescent biosensor revisited. ASM News 62:297–301.
33. Stewart, G. S. A. B., S. A. A. Jassim, S. P. Denyer, P. Newby, K. Linley, and V. K. Dhir. 1998. The specific and sensitive detection of bacterial pathogens within 4 h using bacteriophage amplification. J. Appl. Microbiol. 84:777–783.
34. Ulitzur, S., and J. Kuhn. 1987. Introduction of lux genes into bacteria: a new approach for specific determination of bacteria and their antibiotic susceptibility, PP. 463–472. *In* R. Slomerich, R. Andreesen, A. Kapp, M. Ernst, and W. G. Woods (eds.), Bioluminescence and Chemiluminescence: New Perspectives. Wiley, New York.
35. Ulitzur, S., and J. Kuhn. 1989. Detection and/or identification of microorganisms in a test sample using bioluminescence or other exogenous genetically introduced marker. US Patent 4,861,709.
36. Wilson, S. M., Z. Al Suwaidi, R. McNerney, J. Porter, and F. Drobniewski. 1997. Evaluation of a new rapid bacteriophage-based method for the drug susceptibility testing of *Mycobacterium tuberculosis.* Nat. Med. 3:465–468.
37. Wilson, S. M. 1999. Analytical method using multiple virus labelling. PCT Patent WO99/63348.
38. Wolber, P. K. 1993. Bacterial ice nucleation. Adv. Microb. Physiol. 34:203–237.
39. Wolber, P. K. and R. L. Green. 1990. Detection of bacteria by transduction of Ice nucleation genes. Trends Biotechnol. 8:276–279.
40. Wu, Y., L. Brovko, and M. W. Griffiths. 2001. Influence of phage population on the phage-mediated bioluminescent adenylate kinase (AK) assay for detection of bacteria. Letts. Appl. Microbiol. 33:311–315.

47

Bacteriophages in Bacterial Pathogenesis

PATRICK L. WAGNER
MATTHEW K. WALDOR

Bacteriophages contribute to the virulence of many bacterial pathogens, largely by encoding the structural genes for virulence factors. The most widely recognized phage-encoded virulence factors are bacterial exotoxins, which account for the characteristic clinical manifestations of a number of human diseases caused by bacterial infections. The previous edition of this chapter focused primarily on phage-encoded toxins, but in the intervening years two themes have emerged (169). First, phages are increasingly recognized for encoding genes that contribute to other aspects of bacterial pathogenesis in addition to toxin production. In fact, phage-encoded gene products contribute to virtually every facet of bacterial pathogenicity, from attachment and invasion to immune evasion and transmission among humans. Second, while phages are like other mobile genetic elements in that they disseminate virulence genes among bacterial populations, phages have unique properties that enable them to contribute to bacterial pathogenesis by mechanisms other than transduction as well. For example, virion particles may themselves contain pathogenic components (12, 13); and prophage induction, through gene amplification, transcriptional upregulation, and phage-mediated lysis, can contribute to production and release of virulence factors from bacterial cells (167, 168). Owing to these developments, the contribution of bacteriophages to bacterial pathogenesis can no longer be conceived of simply as transduction of toxin genes that are regulated by the host bacterium. This chapter presents a summary of bacteriophage involvement in bacterial pathogenesis, in which we consider (i) the nature of the bacterial virulence properties altered by phages, (ii) the basic mechanisms by which phages alter these properties, and (iii) the regulation of the phage-related virulence property by the phage and/or its host bacterium.

Bacteriophages in the Pathogenesis of Gram-Positive Organisms

Clostridium botulinum

Clostridium botulinum is a Gram-positive, anaerobic, spore-forming bacillus responsible for botulism. This illness, characterized by flaccid paralysis, is acquired by ingestion of the preformed botulinum toxins, of which there are eight serologically distinct types (A, B, C_1, C_2, D, E, F, and G). Each type consists of a neurotoxin (BoNT) and a hemagglutinin (HA), as well as other nontoxic, nonhemagglutinin (NTNH) components (84). BoNTs are zinc proteases that act inside presynaptic cholinergic neurons at the neuromuscular junction to cleave proteins necessary for fusion of acetylcholine-containing vesicles with the cell membrane, including synaptobrevin, SNAP-25, and syntaxin (117). Each BoNT type specifically cleaves one or more of these vesicle fusion proteins, impairing acetylcholine release and thereby causing paralysis (117).

The genes for the BoNT, HA, and NTNH toxin components are located within clusters that, in types C and D, are located on bacteriophages. The association of botulinum toxin with phages was first described in the 1970s, when nontoxigenic *C. botulinum* cells were rendered toxigenic following exposure to mitomycin-C-induced, cell-free lysates of toxigenic cultures (49, 79, 80). A toxin-converting phage, CEβ, was isolated from such lysates (49), and the toxin structural gene was identified in the CEβ genome (57). Among the botulinum toxin types, only C and D have been shown to be phage-encoded; the remainder are either plasmid-encoded (type G) or assumed to be chromosomal (types A, B, E, and F) (84).

Regulation of BoNT production is poorly understood. A DNA-binding protein, BotR, encoded by a gene (*botR*) located in the toxin gene complex, was shown to increase toxin production when overproduced in *C. botulinum* (106). This protein bound specifically to the promoter regions involved in transcription of the NTNH, BoNT, and HA genes. Since *botR* is adjacent to the toxin genes, it is likely a phage-encoded factor in types C and D, but whether it plays a role in the transcriptional regulation of phage genes other than the toxin genes is unknown. Though early studies identified phage-related differences in the quantity of toxin production (66), the possibility of a relationship between phage transcriptional regulation and botulinum toxin production remains to be explored.

Corynebacterium diphtheriae

Corynebacterium diphtheriae is a Gram-positive bacillus that infects the upper respiratory tract and causes inflammation, necrosis, and the formation of an adherent pseudomembrane. The organism elaborates the potent diphtheria toxin (DT), which is absorbed into the circulation and can cause a systemic, potentially lethal syndrome characterized by muscle weakness or paralysis and circulatory collapse due to myocardial dysfunction. DT is synthesized as a single, inactive polypeptide chain that is cleaved into two functionally distinct fragments: a C-terminal fragment, responsible for entry of toxin into the eukaryotic cell, and an N-terminal fragment, which catalyzes the ADP-ribosylation of eukaryotic elongation factor 2 (EF2) (40).

Phage conversion of DT production by *C. diphtheriae* was discovered in 1951 by Freeman, who demonstrated that stable lysogeny of *C. diphtheriae* with a phage, designated β, was associated with virulence in a guinea pig intoxication model (55). It was subsequently proposed that toxin conversion by β could be due to transduction (7), a hypothesis that was proven by studying recombination between phage β and a related nontoxigenic phage, γ, with a host range different from that of β. Since recombinant phages possessing the host range properties of γ and the toxin-determining property of β could be isolated, toxigenicity was proven to be a discrete genetic attribute of β (61). In 1969, the *tox* locus of the phage was mapped (71), and definitive proof that the structural genes for the toxin were located on the phage genome was finally obtained in 1971, when nitrosoguanidine-induced phage mutants that produced serologically reactive but truncated, nontoxic DT were described (155). The *tox* gene is located near one end of the phage genome, adjacent to the attachment site (71, 96, 109, 144, 145), leading to the hypothesis that the gene was acquired by the phage during an aberrant excision event (96, 129).

Regulation of toxin production by *C. diphtheria* was studied as early as the 1930s, when the negative correlation between environmental iron concentration and DT production by *C. diphtheriae* was described (101, 130). The molecular mechanism of the regulation of DT production by iron concentration has been elucidated more recently. The iron-dependent repressor DtxR, encoded on the *C. diphtheriae* chromosome and postulated to be a global low-iron-response regulator, was shown to repress expression from the diphtheria toxin promoter (25, 140) by directly binding to an adjacent operator (141). Barksdale et al. discovered that UV light, a potent phage-inducing agent, greatly enhanced diphtheria toxin production, but only at low iron levels (6). This result suggested that prophage induction and replication could amplify DT production, but that *tox* transcription relies on the iron concentration being low enough that the DtxR repressor is inactive. These authors' conclusion that DT production depends upon activation of the latent DT-encoding prophage seems unlikely since DT production occurs during all phases of the phage life cycle (59, 107, 108). While these latter observations argue against the necessity of prophage induction for DT production, they do not negate the possibility that prophage induction may contribute to toxin production. It is possible that regulation of DT production integrates environmental stimuli through at least two pathways: prophage induction and the DtxR repression system. The importance of these factors to DT production by *C. diphtheriae* in humans remains relatively unexplored.

Staphylococcus aureus

Staphylococci are Gram-positive cocci that cause a wide range of diseases, including endocarditis, pneumonia, suppurative infections such as skin abscesses and septic arthritis, and diseases attributable to exotoxin production. Suppurative disease is characterized by formation of abscesses, composed largely of staphylococci and neutrophils, which are attracted by staphylococcal components. The nonsuppurative staphylococcal diseases include food poisoning, caused by the ingestion of preformed enterotoxins, as well as toxic shock syndrome (TSS), an acute, potentially lethal illness marked by fever, rash, desquamation, and hypotension that is caused by the toxic shock syndrome toxin (TSST). Many virulence factors produced by *S. aureus* have been isolated and characterized, although it has generally been difficult to establish the role of individual factors (47). Nearly all staphylococci produce degradative enzymes that damage human tissues, and some strains produce a nonenzymatic fibrinolysin that acts as an anticoagulant via activation of plasminogen. Some strains produce toxic exoproteins, including hemolysins, the staphylococcal enterotoxins, the exfoliative toxins, the Panton-Valentine leukocidin (PVL), and TSST. A number of these virulence factors, including the fibrinolysin and most of the exotoxins, are associated with staphylococcal phages.

The staphylococcal exfoliative toxins are responsible for the staphylococcal scalded skin syndrome, which primarily occurs in children. There are two distinct forms of exfoliative toxin, ETA and ETB, which are not directly cytotoxic but rather cause lysis of intercellular attachments in the epidermis, with subsequent blister formation. The gene encoding ETA was thought to be chromosomal until recently, when an ETA-encoding phage, ϕETA, was isolated from *S. aureus*, and a number of clinical *S. aureus* isolates were found to have phage gene sequences near the toxin gene (*eta*) (179). The complete ϕETA genome was sequenced, and the *eta* gene was found near the *att* site of the phage (179). The gene for ETB resides on a family of plasmids (23).

PVL, produced by most strains of *S. aureus*, specifically attacks neutrophils and macrophages (23). The toxin is composed of two protein components that are produced and secreted separately but act synergistically to cause lysis of leukocytes, possibly by altering their permeability to cations. Phage conversion of PVL production was demonstrated as early as 1972 (158), and analysis of sequences adjacent to the PVL genes ultimately led to the isolation of a PVL-encoding phage, named ϕPVL (86, 87). At least one other leukocidin-encoding phage has been isolated from a clinical *S. aureus* isolate (120).

Although best known for causing vomiting during food poisoning, the staphylococcal enterotoxins have also been shown to have superantigen properties (see below), and have been implicated in TSS. At least one staphylococcal enterotoxin gene — *sea* or *entA* (enterotoxin A) — is phage-encoded (17, 32, 39). The enterotoxin E (*entE*) gene was cloned and sequenced; although no *entE*-encoding phage could be detected, the gene was found to be amplified in the presence of UV light and mitomycin C, suggesting that it too could be encoded by an inducible prophage (41). The enterotoxin B (*entB*) gene was identified on a discrete 26.8 kb chromosomal element that has not been further characterized (83).

TSST, consisting of a single 22 kDa polypeptide chain, is one of a number of related toxins capable of producing TSS. The most serious consequences of TSS result from systemic shock, which TSST may produce by one or more of three proposed mechanisms: superantigenicity, enhancement of endotoxin activity, or direct activity on endothelial cells (47). The staphylococcal TSST gene, *tstH*, is present on the bacterial chromosome within a 15.2 kb mobile genetic element designated staphylococcal pathogenicity island 1 (SaPI1) (100). SaPI1 is mobilized at high frequency by the generalized staphylococcal transducing phage 80*α* and depends upon 80α for excision, replication, and encapsidation into 80 α-like phage particles, a relationship reminiscent of the interaction between coliphages P4 and P2 (100) (see chapter 26).

Some staphylococci produce a recently discovered chemotaxis inhibitory protein (CHIPS) that binds to and attenuates the activity of the neutrophil receptors for complement and formylated peptides (160, 161). This function is proposed to protect *S. aureus* from neutrophil-mediated killing, an important host defense against staphylococci. The gene for CHIPS (*chp*) has been shown to reside on a functional quadruple-converting phage that, in addition to transducing the *chp* gene, also transduces the staphylokinase (*sak*) and enterotoxin A (*sea*) genes, and eliminates β-hemolysin production (160). The latter effect presumably occurs via insertional inactivation of the *hlb* gene as is also achieved by lysogenization with another phage, ϕ13 (37, 38). Staphylokinase, or fibrinolysin, is a potent nonenzymatic catalyst of the conversion of plasminogen to plasmin, a clot-dissolving protease. The genes for staphylokinase were cloned from a phage genome (139), and production of staphylokinase was previously shown to be associated with lysogenic conversion by phages other than the CHIPS-encoding phage (32, 92, 174).

Regulation of phage-encoded virulence factors in *S. aureus* is poorly characterized, although a number of phage-associated staphylococcal virulence genes, including *tst*, *entA*, and *eta*, are regulated by chromosomal loci that coordinate expression of catabolic and secreted proteins at particular stages of growth in laboratory culture. These loci include *agr*, active during the shift to post-exponential growth phase, as well as *sar* and *sae*, each of which affects the expression of multiple exoprotein genes (125). Additionally, several environmental signals affect the expression of virulence genes independently of these loci (34, 125). The role of phages in regulating staphylococcal virulence factor production has not been extensively studied, although enterotoxin A production varied with different encoding phages, suggesting that phage factors or processes distinct from the *sea* gene and its associated promoter could influence toxin production (24).

A major obstacle in the treatment of staphylococcal disease is the tendency of staphylococci to develop resistance to antibiotics, generally via alteration of the target molecule of the antibiotic or via acquisition of genes for antibiotic efflux or inactivation (133). Acquisition of antibiotic resistance genes by staphylococci occurs primarily via transduction and conjugation (53). Although no examples of phage-encoded antibiotic resistance genes are known, phages may play an important role in the mobilization of resistance plasmids via generalized transduction and via a poorly understood process termed phage-mediated conjugation, which is mechanistically distinct from transduction (95, 103).

Streptococci

Group A Streptococci

Streptococci are Gram-positive cocci responsible for a number of human diseases, including streptococcal pharyngitis, scarlet fever, impetigo, and bacterial endocarditis. In addition, group A streptococci (GAS) infections can lead, via poorly understood mechanisms, to post-infectious syndromes including acute rheumatic fever, rheumatic heart disease, and acute glomerulonephritis. Like that of staphylococci, the pathogenicity of streptococci is multifactorial, involving both bacterial structural components and exoproteins elaborated by the bacteria. Since lysogeny is very common among GAS (113) and transduction is the only known naturally occurring mechanism of horizontal gene exchange among then, it has been proposed that phages play an important role in periodic genetic shifts in GAS that may explain variation in the clinical features of infection (30, 90, 113). Furthermore, a number of specific virulence properties of GAS are influenced by bacteriophages.

The streptococcal pyrogenic exotoxins (SPEs) are related proteins (SpeA, SpeB, SpeC) released by GAS that cause fever and toxic shock syndrome, and enhance the activity of endotoxin (111). These toxins also lead to the characteristic scarlet fever rash, mediated by delayed-type hypersensitivity in the skin and nonspecific T cell stimulation (superantigenicity) (111). As early as 1927, Frobisher and Brown described experiments in which nontoxigenic streptococcal strains acquired the ability to produce SPE following exposure to filtered supernatants of cultures of toxigenic streptococci (56). Much later, this conversion was demonstrated to be mediated by an SpeA-encoding bacteriophage (181). The structural genes for SpeA (85, 172) and SpeC (60) were subsequently found to be located on the genomes of temperate phages, while SpeB is believed to be encoded by the streptococcal chromosome (113). There is evidence that toxin-encoding phages may play a direct role in regulating SpeA and SpeC production. Zabriskie (181) demonstrated enhanced phage and streptococcal pyrogenic exotoxin A (scarlatinal toxin) production by GAS in response to UV light (181), suggesting that prophage induction could contribute to SpeA production. Recently, Broudy et al. discovered a soluble phage-inducing factor (SPIF), elaborated by human pharyngeal epithelial cells, that induced an SpeC-encoding phage (60) and resulted in increased toxin production by GAS (26). Although the identity of the SPIF molecule and the mechanism by which it induces phage and toxin production remain to be explored, these findings imply a role for prophage induction in toxin production by GAS.

Many strains of streptococci also produce hyaluronidases, enzymes capable of hydrolyzing hyaluronic acid, a component both of the bacterial capsule and of human connective tissue. In some cases, hyaluronidase genes are located on phage genomes (77), and hyaluronidase is incorporated into the virion particles (10, 11), possibly aiding the phage in capsule penetration during infection of or release from the streptococcal cell. Whether the phage-associated hyaluronidase activity, including that of the virion particles themselves, aids the bacterium in its spread through or destruction of human connective tissues is unknown, although antibody to phage-encoded hyaluronidase is detectable in the serum of patients with GAS infections (63).

The M protein, which is a major cell surface antigen of GAS and is its principal virulence factor, confers resistance to phagocytosis (54). Growth under conditions known to cure bacteria of plasmids or phages led to a reduction in M protein expression by GAS (36), suggesting that a mobile genetic element could be involved in M protein regulation. Subsequently, Spanier and Cleary described a temperate bacteriophage, SP24, that, upon lysogenization, greatly enhanced M protein production by streptococci (149). These investigators proposed that a hypothetical phage-encoded gene, *mprA*, upregulates expression of the chromosomal gene encoding the M protein.

Temperate GAS phages have also been observed to play a role in the transfer of antibiotic resistance through transduction of resistance genes (30, 76, 154), although as in staphylococci there is no evidence that resistance genes are phage-encoded. Similarly, streptolysin S, a β-hemolysin with lytic activity against erythrocytes and leukocytes, was shown to be transduced by a derivative of phage A25 (147), although subsequent studies have identified a chromosomal locus responsible for streptolysin S synthesis (124).

Streptococcus mitis

The viridans streptococci, including *Streptococcus mutans* and *Streptococcus mitis*, colonize the oropharynx and can cause infectious endocarditis after gaining access to the bloodstream. A critical early step in the pathogenesis of infectious endocarditis is bacterial adhesion to the endocardium, platelets, and fibrin, which together form the characteristic vegetations on affected heart valves. Two *S. mitis* proteins, encoded by an inducible prophage (SM1), are found in the *S. mitis* membrane and are important in adhesion of *S. mitis* to platelets (12, 13). These proteins, PblA and PblB, resemble phage capsid components and are constituents of SM1 phage particles (13). It remains unknown whether the membrane-bound or the virion-associated forms of PblA and PblB, or both, mediate bacterial adhesion to platelets. UV light was shown to increase expression of the platelet-binding loci of *S. mitis* (12, 13), raising the possibility that prophage induction, perhaps by agents in the blood or endocardium, plays a role in expression of PblA and PblB in vivo.

Streptococcus pneumoniae

Streptococcus pneumoniae is a common cause of pneumonia, otitis media, meningitis, and septicemia. One of the many factors contributing to the virulence of this organism is a cell wall degradative enzyme, autolysin (15, 16). The pathogenic mechanisms of autolysin are not well characterized, but may involve release of pneumococcal virulence factors or structural components. The majority of pneumococcal clinical isolates contain multiple autolysin-encoding loci, and many pneumococcal prophage genomes hybridize with an autolysin gene (*lytA*) probe, suggesting that autolysins may actually be phage lysis proteins (134). In fact, a recently sequenced temperate pneumococcal phage encodes a gene with approximately 90% identity to *lytA* (58). Phages may thus contribute to the as yet poorly defined role of autolysin in pneumococcal virulence.

Bacteriophages in the Pathogenesis of Gram-Negative Organisms

Bordetella avium

Bordetella avium, which is genetically related to *B. pertussis*, causes avian bordetellosis, a highly morbid pertussis-like illness in young turkeys (146). *B. avium* shares with *B. pertussis* several virulence factors, including dermonecrotic toxin, tracheal cytotoxin, hemagglutinin, and fimbriae. *B. avium* also encodes pertussis toxin (PT). Unlike in *B. pertussis*, however, PT in *B. avium* is encoded by a temperate generalized transducing phage, Ba1, although *B. avium* has not been shown to produce measurable amounts of PT (142). It was speculated that PT could be produced as a late phage gene product and released via phage-mediated lysis (159), although mitomycin C treatment does not induce the Ba1 prophage or PT production by *B. avium* (152), and there is no evidence to date that Ba1 plays a role in *B. avium* pathogenesis (143).

Escherichia coli

Escherichia coli, a normal inhabitant of the human gastrointestinal tract, causes a variety of human diseases including urinary tract infections, pneumonia, meningitis, septicemia, and a wide range of gastrointestinal diseases. The numerous virulence factors of *E. coli* range from structural features (endotoxin, capsules, fimbriae, and adhesins) to exotoxins, which are particularly important in gastrointestinal disease and include the heat-stable and heat-labile enterotoxins. A subset of diarrheogenic *E. coli*, including *E. coli* O157:H7, produce a group of related toxins called Shiga toxins (Stx). These toxins cause severe clinical manifestations in addition to diarrhea (22, 127, 164), including hemorrhagic colitis and the hemolytic uremic syndrome (HUS), a potentially fatal systemic syndrome characterized by acute renal failure, microangiopathic hemolytic anemia, and thrombocytopenia. There are two major classes of Stx (Stx1 and Stx2), each of which is composed of a pentameric B subunit involved in transport into the eukaryotic cell, and a monomeric enzymatic A subunit that cleaves 28S rRNA (1), leading to inhibition of protein synthesis.

Stx production by *E. coli* provides the most compelling known example of virulence factor regulation by the life cycle of a bacteriophage. The Shiga toxin genes are, in all known cases, closely associated with bacteriophage sequences, and are often present on functional lambdoid prophages (32, 42, 89, 116, 122, 156, 166). Mitomycin C and other phage-inducing antibiotics greatly enhance toxin production by shiga-toxin-producing *E. coli* (STEC) (2, 3, 69, 74, 75), and a mechanism for this relationship was suggested by genome sequence analysis of several Stx-encoding phages, including H-19B (Stx1) and 933W (Stx2), which revealed that the *stx* genes are located within the tightly regulated late operon (105, 115, 121, 132, 180). Thus, the toxin genes could be regulated as late phage genes, transcribed in concert with the phage lysis and morphogenesis genes at the appropriate time following prophage induction.

Recently, the importance of prophage induction in Stx production and release has been established, as have the mechanisms underlying this relationship. Stx2 production by lysogens of phage ϕ361 relied almost exclusively on the late phage promoter ($p_R{}'$) and its associated antiterminator gene (*Q*), both in vitro and in a mouse intestinal infection model (168), indicating that Stx2 production ultimately depends upon prophage induction. In contrast, in lysogens of phage H-19B, late phage transcription was not essential for Stx1 production (167), due to the presence of an iron-regulated promoter (p_{Stx1}) adjacent to the stx_1 coding sequence (28, 29, 44, 81). Nevertheless, phage replication and phage-regulated transcription initiating at the p_R and $p_R{}'$ promoters each contributed to the increase in Stx1 production observed following prophage induction (167). Moreover, phage-mediated lysis controlled the duration and therefore the total amount of Stx1 production, and allowed for Stx1 release from the cell (167). The central role of prophage induction in Stx production and release could explain why prophage-inducing antibiotics, often used in the treatment of STEC infection, are epidemiologically associated with adverse clinical outcomes (27, 31, 91, 131, 175). Furthermore, human cells (neutrophils) and molecules released from them (H_2O_2) can induce Stx-encoding phages and subsequent Stx production (165), raising the possibility that agents present in the human body may induce Stx-encoding prophages and thereby contribute to STEC pathogenicity.

A number of other coliphages encode factors known or suspected to have a role in *E. coli* pathogenesis, including phage λ, which has long been known to enhance the serum survival of its *E. coli* lysogens (118). Barondess and Beckwith identified a novel outer membrane lipoprotein (Bor) that is encoded by λ, secreted during lysogeny, and important for serum resistance of λ lysogens (8, 9). A second λ-encoded outer membrane protein (Lom) (8, 135) may contribute to the adhesive properties of *E. coli*, since the reported advantage of λ lysogens observed in adhesion to buccal epithelial cells was absent in a λ lysogen bearing a *lom*::Tn*phoA* fusion (128). The contribution of *lom* and *bor* to *E. coli* pathogenesis has not been tested in animal models of infection. λ -like phages isolated from pathogenic *E. coli* have also been shown to encode a hemolysin (Ehly) (18–20, 151), although the significance of Ehly as a virulence factor is uncertain.

Pseudomonas aeruginosa

Pseudomonas aeruginosa is an important opportunistic pathogen, commonly responsible for pneumonia and infections of burns, wounds, and the urinary tract. Among the many factors contributing to the virulence of *P. aeruginosa* is a pore-forming cytotoxin (CTX) (4, 21, 126) that confers virulence in an animal model (5) and is encoded by a phage, ϕCTX (67). In this phage, the *cos* site lies between the nearby toxin gene and the *attP* site, implying that a more complicated mechanism than aberrant excision must account for the acquisition of the toxin gene by ϕCTX (67, 68, 119). Little is known regarding the control of CTX production; however, Xiong et al. suggested that ϕCTX integration into the chromosome, which brings an active promoter into continuity with the *ctx* genes, could be critical for CTX production (178). Consistent with this hypothesis, CTX production in cultures of ϕCTX-infected *P. aeruginosa* temporally followed ϕCTX integration (178). This requirement would make ϕCTX unusual among toxin-encoding phages, as there are no other cases in which phage integration is known to upregulate toxin gene transcription.

Other phages may contribute to the pathogenesis of *P. aeruginosa* in other ways. Several phages alter the chemical composition and structure of the O-antigen of *P. aeruginosa* upon lysogenization (14, 33, 46, 48, 70, 94, 97, 102, 104). These changes alter bacterial susceptibility to phage superinfection (70), and in principle could aid the organism in evading immune defenses (136). Another *P. aeruginosa* phage, FIZ15, has been shown to enhance adhesion of the organism to buccal epithelial cells (157).

Salmonella enterica

Salmonella species are responsible for a number of clinical syndromes, the most important of which are gastroenteritis and typhoid fever; the pathogenesis of the latter involves a number of phage-related virulence properties. Certain *Salmonella* species, including *S. typhi*, penetrate the intestinal epithelial barrier and subsequently invade the subepithelial tissues and regional lymphatic structures, including Peyer's patches. Following invasion, these *Salmonella* species have the ability to survive within phagocytes and spread hematogenously to involve lymphoid structures throughout the body. Subsequently, organisms can reinfect the intestinal tract, especially the gall bladder, and can cause late complications, including intestinal perforation, hemorrhage, and abscess formation. Invasion by *Salmonella* requires a type III secretion system encoded by *Salmonella* pathogenicity island 1 (SPI1), which injects effector proteins directly into the cytoplasm of human host cells (73). SopE, a protein that activates human Rho GTPases and facilitates entry of *Salmonella* into tissue culture cells, is one such effector protein (64, 65, 176), and is encoded on SopEϕ, a P2-like phage (114) (P2-like bacteriophages are reviewed in chapter 25). This is a remarkable instance of two mobile genetic elements (SPI1 and SopEϕ) cooperating to enhance the virulence of *Salmonella*.

Following translocation of the intestinal epithelium, *S. typhimurium* preferentially localizes to Peyer's patches, where a gene (*gipA*) encoded in the late operon of another *Salmonella* phage, Gifsy-1, is specifically upregulated (150). A *gipA* null mutant was selectively impaired for growth in the Peyer's patch following oral inoculation, but not impaired in general growth, attachment, invasion, or virulence following intraperitoneal injection (150). The O-antigen is thought to be an important virulence determinant of *Salmonella*, determining its susceptibility to complement and other serum proteins, and to phagocytosis (136). In *Salmonella*, the O-antigen structure and composition are altered by phage-encoded enzymes (137, 138, 177). Since the structure of the O-antigen is one determinant of the degree of susceptibility to phagocytes and serum factors (82, 99), O-antigen alteration by bacteriophages could be an important determinant of *Salmonella* virulence.

Inside the phagocyte, *Salmonella* is subjected to oxidative stress within eukaryotic organelles that produce reactive oxygen intermediates, including the superoxide radical. *Salmonella* resists oxidative damage using enzymes such as superoxide dismutase, which catalyzes the conversion of superoxide ion into hydrogen peroxide and molecular oxygen and has been implicated in *Salmonella* pathogenesis (45, 50). Figueroa-Bossi and Bossi showed that a *Salmonella* superoxide dismutase, SodC, is encoded by Gifsy-2, a functional bacteriophage capable of transducing *sodC* (52). Curing *S. typhimurium* of this phage resulted in attenuated virulence in a mouse infection model (52). Hydrogen peroxide was a highly effective inducer of Gifsy-2 (52), suggesting a possible relationship between SodC activity (which results in hydrogen peroxide production) and Gifsy-2 induction.

Shigella

Shigella species cause dysentery, a disease characterized by abdominal pain, tenesmus, and bloody diarrhea. *Shigella* organisms attach to and penetrate colonic epithelial cells, inside which they multiply and spread to adjacent cells, leading to inflammation, ulceration, and frequent discharge of stools containing blood, mucus, and pus. Two virulence factors of *S. dysenteriae* are affected by phages: the O-antigen and Shiga toxin. The O-antigen likely contributes to bacterial evasion of human immune defenses by conferring resistance to complement activation and phagocytosis (136). Phage-encoded enzymes alter the O-antigen, as in *P. aeruginosa* and *Salmonella* (above), which presumably benefits the phage by conferring superinfection resistance on the host bacterial cell (62, 110). However, since O-antigen alteration is a mechanism of immune evasion by bacteria, these phages may also make an indirect contribution to the *Shigella* pathogenicity (136).

Shiga toxin may damage intestinal epithelial cells by inhibiting protein synthesis in a manner identical to that of the Shiga toxins of *E. coli* (above). McDonough and Butterton sequenced an approximately 32 kb segment of the S. dysenteriae type 1 chromosome and found that *stxAB*, long considered to be chromosomal genes, are embedded within phage-like sequences, including a gene with more than 95% homology with the *S* lysis gene of phage 933W located 3′ relative to *stxAB* (112). A large number of insertion sequences were also located near the *stx* genes, implying that they are present on a prophage rendered defective by the insertion sequences.

Vibrio cholerae and Vibrio parahaemolyticus

The curved Gram-negative rod *Vibrio cholerae* causes cholera, a severe, sometimes lethal, diarrheal disease that often occurs in epidemics. *V. cholerae* colonizes the small bowel and elaborates cholera toxin (CT), which accounts for the profuse watery diarrhea characteristic of cholera. CT consists of a pentameric B subunit, which binds to a receptor on epithelial cells, and a monomeric enzymatic A subunit, which transfers ADP-ribose to a G protein that regulates the activity of adenylate cyclase (171). The resulting increase in intracellular cAMP concentration leads to voluminous secretion of chloride and water from epithelial cells into the small bowel lumen. *V. cholerae* colonization of the small intestine requires TCP, a bundle-forming pilus, the production of which, like CT, requires the transcriptional activator, ToxR (148).

CT is encoded within the genome of CTXϕ, the first filamentous bacteriophage shown to participate in the lysogenic conversion of its host bacterium (170). Unlike the well-characterized F-pilus-specific filamentous coliphage M13 (see chapter 12), CTXϕ integrates into the *V. cholerae* genome. Interestingly, the *V. cholerae* receptor for CTXϕ is TCP, and since TCP is produced during *V. cholerae* intestinal colonization, it has been proposed that CTXϕ infection of *V. cholerae* occurs most efficiently within the human intestine (170). Thus, lysogenic conversion of an ancestral TCP^{+} nonlysogenic *V. cholerae* strain to toxigenicity (creating a fully pathogenic *V. cholerae*) may have occurred within the human host. Transcription of *ctxAB*, which encodes CT, is thought to depend largely on two activators: the chromosome-encoded ToxR and the *Vibrio* pathogenicity island (VPI)-encoded ToxT (148). The interaction between ToxT and the *ctxAB* promoter provides another example of interplay between mobile genetic elements (VPI and CTXϕ) in bacterial pathogenicity, akin to that between SPI1 and SopEϕ in *Salmonella*. VPI has been proposed to correspond to the genome of a prophage, VPIϕ (88), although this hypothesis awaits confirmation. Phage-initiated transcription at a promoter other than the *ctxAB* promoter can also direct *ctxAB* messenger RNA production (98), but the significance of this ToxR- and ToxT-independent *ctxAB* transcription for *V. cholerae* pathogenicity remains unknown.

Before the discovery of CTXϕ, two gene products (Ace and Zot) encoded within the 6.9 kb CTXϕ genome were proposed to have enterotoxic activity. The Ace protein (153) is now thought to be a CTXϕ minor coat protein (170); therefore, CTXϕ secretion from *V. cholerae* may constitute a mechanism for delivery of toxic virion particles to the intestinal epithelium. Zot (51) is an ortholog of the M13 pI protein (93), which is thought to act in phage secretion (43). The contribution to *V. cholerae* pathogenicity of these two enterotoxins, now known to be essential phage gene products, awaits further study.

Vibrio parahaemolyticus, like *V. cholerae*, causes diarrhea, although its virulence factors are not well characterized. Until recently, *V. parahaemolyticus* strains were not thought to give rise to pandemics. However, one *V. parahaemolyticus* serovar, O3:K6, may currently be causing a pandemic of diarrheal disease in North America and Asia (35). The filamentous phage f237, related to CTXϕ but lacking *ctxAB*, is closely associated with these *V. parahaemolyticus* isolates (78), and ongoing studies are aimed at determining whether f237 played a direct role in the emergence of this pathogen.

Bacteriophages in the Pathogenesis of Mycoplasma

Mycoplasma arthritidis

Mycoplasma arthritidis causes arthritis in rodents, and the arthritogenesis of *M. arthritidis* was found to be directly correlated with lysogeny of this organism by a bacteriophage designated MAV1 (163). The approximately 16 kb genome of this phage was sequenced, leading to the identification of an open reading frame, *vir*, that encodes a product with a lipoprotein signal sequence (162). The contribution to *M. arthritidis* pathogenicity of this gene, which is encoded in the opposite orientation relative to all other MAV1 genes and transcribed during lysogeny,

Table 47-1 Contribution of Bacteriophages to the Virulence of Bacterial Pathogens

Bacterial Pathogen	Mechanism	References
Gram-positive-pathogens		
C. botulinum	Botulinum toxin is phage-encoded	57
C. diphtheriae	Diphtheria toxin is phage-encoded	71, 155
S. aureus	Fibrinolysin is phage-encoded	139
	Staphylococcal enterotoxins are phage-encoded	17, 32, 39
	Staphylococcal exfoliative toxins are phage-encoded	179
	Toxic shock syndrome toxin is encoded by SapI1, a mobile pathogenicity island transduced at high frequency by φ80	100
	Generalized transduction contributes to horizontal transmission of staphylococcal antibiotic resistance genes	53
S. pyogenes (GAS)	Hyaluronidase is phage-encoded	77
	Phages encode CHIPS, a phagocytotoxin	160
	Phage φPVL encodes the Panton-Valentine leukocidin	87
	Lysogeny upregulates the antiphagocytic M protein	149
	Streptococcal pyrogenic (erythrogenic, scarlatinal) exotoxins are phage-encoded	60, 85, 172
	Generalized transduction contributes to horizontal transmission of streptococcal antibiotic resistance genes	30
S. pneumoniae	Phages encode autolysins	58, 134
S. mitis	The SM1-encoded PblA and PblB surface proteins promote adhesion to platelets	12, 13
P. aeruginosa	Phage FIZ15 promotes adhesion to buccal epithlial cells	157
V. cholerae	The toxin-coregulated pilus may be phage-encoded	88
Gram-negative-pathogens		
B. avium	Pertussis toxin is phage-encoded in *B. avium*	159
E. coli	The λ-encoded *bor* gene confers a survival advantage in animal serum	8
	The λ -encoded *lom* gene promotes adhesion to buccal epithelial cells	8, 128, 135
	The Shiga toxins are phage-encoded and their production is regulated via prophage induction	72, 123, 167, 168, 173
P. aeruginosa	Phages are associated with O-antigen variation	70
	Pseudomonas cytotoxins are phage-encoded	67
S. enterica	Phage SopEφ transduces a type III secretion system effector that promotes entry into epithelial cells	114
	Phage Gifsy-1 encodes *gipA*, a gene that enhances survival in the Peyer's patches	150
	Phage Gifsy-2 encodes SodC, a superoxide dismutase	52
	Phages are associated with O-antigen variation	137, 138, 177
S. dysenteriae	Phages are associated with O-antigen variation	62, 110
	The Shiga toxin genes are associated with phage sequences, probably a defective prophage	112
V. cholerae	Cholera toxin is phage-encoded	170
	The Ace and Zot enterotoxins are essential phage proteins	170
V. parahaemolyticus	The pandemic serovar O3:K6 is associated with phage f237	78
Mycoplasma		
M. arthritidis	The phage MAV1-encoded *vir* gene may contribute to arthritogenesis	162, 163

remains to be explored. MAV1 and other mycoplasma phages are reviewed in chapter 40.

Conclusions

Bacteriophages play an important role in bacterial pathogenesis through a variety of mechanisms (table 47-1). They promote pathogen evolution by serving as vectors for the dissemination of virulence genes. Some phages regulate, through the process of prophage induction, when and where these virulence genes are expressed. Phage particles may contain structural components that act in pathogenesis, and some phages alter bacterial cell surface antigens in ways that could contribute to bacterial evasion of human immune defenses. Understanding new ways in which bacteriophages contribute to bacterial pathogenesis is a fertile area for future research, and such knowledge could suggest novel strategies for the prevention and treatment of bacterial infections.

References

1. Acheson, D.W. K., A. Donohue-Rolfe, and G. T. Keusch. 1991. The family of Shiga and Shiga-like toxins, pp. 415–433. *In* J. E. Alouf, and J. H. Freer (eds.) Sourcebook of Bacterial Protein Toxins. Academic Press, London.
2. Acheson, D.W. K., A.V. Kane, G. T. Keusch, and A. Donohue-Rolfe. 1989. High yield purification and subunit characterization of Shiga-like toxin II, abstract #901. 29th Interscience Conference on Antimicrobial Agents and Chemotherapy.
3. Al-Jumaili, I., D. A. Burke, S. M. Scotland, H. M. Al-Mardini, and C. O. Record. 1992. A method of enhancing verocytotoxin production by *Escherichia coli.* FEMS Microbiol. Lett. 93:121–126.
4. Baltch, A. L., M. C. Hammer, R. P. Smith, T. G. Obrig, J.V. Conroy, M. B. Bishop, M. A. Egy, and F. Lutz. 1985. Effects of *Pseudomonas aeruginosa* cytotoxin on human serum and granulocytes and their microbicidal, phagocytic, and chemotactic functions. Infect. Immun. 48:498–506.
5. Baltch, A. L., R. P. Smith, M. Franke, W. Ritz, P. Michelsen, L. Bopp, and F. Lutz. 1994. *Pseudomonas aeruginosa* cytotoxin as a pathogenicity factor in a systemic infection of leukopenic mice. Toxicon 32:27–34.
6. Barksdale, L., L. Garmise, and K. Horibata. 1960. Virulence, toxinogeny and lysogeny in *Corynebacterium diphtheriae.* Ann. N.Y. Acad. Sci. 88:1093–1098.
7. Bardsdale, W. L., and A. M. Pappenheimer. 1954. Phage–host relationships in nontoxigenic and toxigenic diphtheria bacilli. J. Bacteriol. 67:220–232.
8. Barondess, J. J., and J. Beckwith. 1990. A bacterial virulence determinant encoded by lysogenic coliphage lambda. Nature 346:871–874.
9. Barondess, J. J., and J. Beckwith. 1995. *bor* gene of phage lambda, involved in serum resistance, encodes a widely conserved outer membrane lipoprotein. J. Bacteriol. 177:1247–1253.
10. Benchetrit, L. C., E. D. Gray, R. D. Edstrom, and L. W. Wannamaker. 1978. Purification and characterization of a hyaluronidase associated with a temperate bacteriophage of group A, type 49 streptococci. J. Bacteriol. 134:221–228.
11. Benchetrit, L. C., E. D. Gray, and L. W. Wannamaker. 1977. Hyaluronidase activity of bacteriophages of group A streptococci. Infect. Immun. 15:527–532.
12. Bensing, B. A., C. E. Rubens, and P. M. Sullam. 2001. Genetic loci of *Streptococcus mitis* that mediate binding to human platelets. Infect. Immun. 69:1373–1380.
13. Bensing, B. A., I. R. Siboo, and P. M. Sullam. 2001. Proteins PblA and PblB of *Streptococcus mitis,* which promote binding to human platelets, are encoded within a lysogenic bacteriophage. Infect. Immun. 69:6186–6192.
14. Bergan, T., and T. Midtvedt. 1975. Epidemiological markers for *Pseudomonas aeruginosa.* Acta Pathol. Microbiol. Scand. [B] 83:1–9.
15. Berry, A. M., R. A. Lock, D. Hansman, and J. C. Paton. 1989. Contribution of autolysin to virulence of *Streptococcus pneumoniae.* Infect. Immun. 57:2324–2330.
16. Berry, A. M., and J. C. Paton. 2000. Additive attenuation of virulence of *Streptococcus pneumoniae* by mutation of the genes encoding pneumolysin and other putative pneumococcal virulence proteins. Infect. Immun. 68:133–140.
17. Betley, M. J., and J. J. Mekalanos. 1985. Staphylococcal enterotoxin A is encoded by phage. Science 229:185–187.
18. Beutin, L., L. Bode, M. Ozel., and R. Stephan. 1990. Enterohemolysin production is associated with a temperate bacteriophage in *Escherichia coli* serogroup O26 strains. J. Bacteriol. 172:6469–6475.
19. Beutin, L., J. Prada, S. Zimmermann, R. Stephan, I. Orskov, and F. Orskov. 1988. Enterohemolysin, a new type of hemolysin produced by some strains of enteropathogenic *E. coli* (EPEC). Zentralbl. Bakteriol. Mikrobiol. Hyg. [A] 267:576–588.
20. Beutin, L., U. H. Stroeher, and P. A. Manning. 1993. Isolation of enterohemolysin (Ehly2)-associated sequences encoded on temperate phages of *Escherichia coli.* Gene 132:95–99.
21. Bishop, M. B., A. L. Baltch, L. A. Hill, R. P. Smith, F. Lutz, and M. Pollack. 1987. The effect of *Pseudomonas aeruginosa* cytotoxin and toxin A on human polymorphonuclear leukocytes. J. Med. Microbiol. 24:315–324.
22. Boerlin, P., S. A. McEwen, F. Boerlin-Petzold, J. B. Wilson, R. P. Johnson, and C. L. Gyles. 1999. Associations between virulence factors of Shiga toxin-producing *Escherichia coli* and disease in humans. J. Clin. Microbiol. 37:497–503.
23. Bohach, G. A., and T. J. Foster. 2000. *Staphylococcus aureus* exotoxins, pp. 367–378. *In* V. A. e. a. Fischetti (ed.) Gram-Positive Pathogens. ASM Press, Washington, D.C.
24. Borst, D. W., and M. J. Betley. 1994. Phage-associated differences in staphylococcal enterotoxin A gene (*sea*) expression correlate with *sea* allele class. Infect. Immun. 62:113–118.

25. Boyd, J., M. N. Oza, and J. R. Murphy. 1990. Molecular cloning and DNA sequence analysis of a diphtheria *tox* iron-dependent regulatory element (*dtxR*) from *Corynebacterium diphtheriae.* Proc. Natl. Acad. Sci. USA 87:5968–5972.
26. Broudy, T. B., V. Pancholi, and V. A. Fischetti. 2001. Induction of lysogenic bacteriophage and phage-associated toxin from group A streptococci during coculture with human pharyngeal cells. Infect. Immun. 69:1440–1443.
27. Butler, T., M. R. Islam, M. A. K. Azad, and P. K. Jones. 1987. Risk factors for development of hemolytic uremic syndrome during shigellosis. J. Pediatr. 110:894–897.
28. Calderwood, S. B., F. Auclair, A. Donohue-Rolfe, G. T. Keusch, and J. J. Mekalanos. 1987. Nucleotide sequence of the Shiga-like toxin genes of *Escherichia coli.* Proc. Natl. Acad. Sci. USA 84:4364–4368.
29. Calderwood, S. B., and J. J. Mekalanos. 1987. Iron regulation of Shiga-like toxin expression in *Escherichia coli* is mediated by the *fur* locus. J. Bacteriol. 169:4759–4764.
30. Caparon, M. 2000. Genetics of group A streptococci, pp. 53–64. *In* V. A. e. a. Fischetti (eds.) Gram-Positive Pathogens. ASM Press, Washington, D.C.
31. Carter, A. O., A. A. Borczyk, J. A. Carlson, B. Harvey, J. C. Hockin, M. A. Karmali, C. Krishnan, D. A. Korn, and H. Lior. 1987. A severe outbreak of *Escherichia coli* O157:H7-associated hemorrhagic colitis in a nursing home. N. Engl. J. Med. 317:1496–1500.
32. Casman, E. P. 1965. Staphylococcal enterotoxin. Ann N. Y. Acad. Sci. 128:124–131.
33. Castillo, F. J., and P. F. Bartell. 1974. Studies on the bacteriophage 2 receptors of *Pseudomonas aeruginosa.* J. Virol. 14:904–909.
34. Chan, P. F., and S. J. Foster. 1998. The role of environmental factors in the regulation of virulence-determinant expression in *Staphylococcus aureus* 8325-4. Microbiology 144:2469–2479.
35. Chowdhury, N. R., S. Chakraborty, T. Ramamurthy, M. Nishibuchi, S. Yamasaki, Y. Takeda, and G. B. Nair. 2000. Molecular evidence of clonal *Vibrio parahaemolyticus* pandemic strains. Emerg. Infect. Dis. 6:631–636.
36. Cleary, P. P., Z. Johnson, and L. W. Wannamaker. 1975. Genetic instability of M protein and serum opacity factor of group A streptococci: evidence suggesting extrachromosomal control. Infect. Immun. 12:109–118.
37. Coleman, D., J. Knights, R. Russell, D. Shanley, T. H. Birkbeck, G. Dougan, and I. Charles. 1991. Insertional inactivation of the *Staphylococcus aureus* beta-toxin by bacteriophage phi 13 occurs by site- and orientation-specific integration of the phi 13 genome. Mol. Microbiol. 5:933–939.
38. Coleman, D. C., J. P. Arbuthnott, H. M. Pomeroy, and T. H. Birkbeck. 1986. Cloning and expression in *Escherichia coli* and *Staphylococcus aureus*: evidence that bacteriophage conversion of beta-lysin activity is caused by insertional inactivation of the beta-lysin determinant. Microb. Pathog. 1:549–564.
39. Coleman, D. C., D. J. Sullivan, R. J. Russell, J. P. Arbuthnott, B. F. Carey, and H. M. Pomeroy. 1989. *Staphylococcus aureus* bacteriophages mediating the simultaneous lysogenic conversion of β-lysin, staphylokinase and enterotoxin A: molecular mechanism of triple conversion. J. Gen. Microbiol. 135:1679–1697.
40. Collier, R. J. 2001. Understanding the mode of action of diphtheria toxin: a perspective on progress during the 20th century. Toxicon 39:1793–1803.
41. Couch, J. L., M. T. Soltis, and M. J. Betley. 1988. Cloning and nucleotide sequence of the type E staphylococcal enterotoxin gene. J. Bacteriol. 170:2954–2960.
42. Datz, M., C. Janetzki-Mittman, S. Franke, F. Gunzer, H. Schmidt, and H. Karch. 1996. Analysis of the enterohemorrhagic *Escherichia coli* O157 DNA region containing lambdoid phage gene *p* and Shiga-like toxin structural genes. Appl. Environ. Microbiol. 62:791–797.
43. Davis, B. M., E. H. Lawson, M. Sandkvist, A. Ali, S. Sozhamannan, and M. K. Waldor. 2000. Convergence of the secretory pathways for cholera toxin and the filamentous phage, CTXphi. Science 288:333–335.
44. De Grandis, S., J. Ginsberg, M. Toone, S. Climie, J. Friesen, and J. Brunton. 1987. Nucleotide sequence and promoter mapping of the *Escherichia coli* Shiga-like toxin operon of bacteriophage H-19B. J. Bacteriol. 169:4313–4319.
45. De Groote, M. A., U. A. Ochsner, M. U. Shiloh, C. Nathan, J. M. McCord, M. C. Dinauer, S. J. Libby, A. Vazquez-Torres, Y. Xu, and F. C. Fang. 1997. Periplasmic superoxide dismutase protects *Salmonella* from products of phagocyte NADPH-oxidase and nitric oxide synthase. Proc. Natl. Acad. Sci. USA 94:13997–14001.
46. Dimitracopoulos, G., and P. F. Bartell. 1979. Phage-related surface modifications of *Pseudomonas aeruginosa*: effects on the biological activity of viable cells. Infect. Immun. 23:87–93.
47. Dinges, M. M., P. M. Orwin, and P. M. Schlievert. 2000. Exotoxins of *Staphylococcus aureus.* Clin. Microbiol. Rev. 13:16–34.
48. Egan, J. B., and B. W. Holloway. 1961. Genetic studies on lysogeny in *Pseudomonas aeruginosa.* Aust. J. Exp. Biol. 39:9–18.
49. Eklund, M., F. Pysky, S. Reed, and C. Smith. 1971. Bacteriophage and the toxigenicity of *Clostridium botulinum* Type C. Science 172:480–482.
50. Farrant, J. L., A. Sansone, J. R. Canvin, M. J. Pallen, P. R. Langford, T. S. Wallis, G. Dougan, and J. S. Kroll. 1997. Bacterial copper- and zinc-cofactored superoxide dismutase contributes to the pathogenesis of systemic salmonellosis. Mol. Microbiol. 25:785–796.
51. Fasano, A., B. Baudry, D. W. Pumplin, S. S. Wasserman, B. D. Tall, J. M. Ketley, and J. B. Kaper. 1991. *Vibrio cholerae* produces a second enterotoxin, which affects intestinal tight junctions. Proc. Natl. Acad. Sci. USA 88:5242–5246.
52. Figueroa-Bossi, N., and L. Bossi. 1999. Inducible prophages contribute to *Salmonella* virulence in mice. Mol. Microbiol. 33:167–176.
53. Firth, N., and R. A. Skurray. 2000. Genetics: accessory elements and genetic exchange, pp. 326–338. *In* V. A. e. a. Fischetti (ed.) Gram-Positive Pathogens. ASM Press, Washington, D.C.

54. Fischetti, V. A., K. F. Jones, S. K. Hollingshead, and J. R. Scott. 1988. Structure, function, and genetics of streptococcal M protein. Rev. Infect. Dis. 10 [Suppl 2]:S356–359.
55. Freeman, V. J. 1951. Studies on the virulence of bacteriophage-infected strains of *Corynebacterium diphtheriae.* J. Bacteriol. 61:675–688.
56. Frobisher, M., and J. Brown. 1927. Transmissible toxicogenicity of streptococci. Bull. Johns Hopkins Hosp. 41:167–173.
57. Fujii, N., K. Oguma, N. Yokosawa, K. Kimura, and K. Tsuzuki. 1988. Characterization of bacteriophage nucleic acids obtained from *Clostridium botulinum* types C and D. Appl. Environ. Microbiol. 54:69–73.
58. Garcia, P., V. Obregon, E. Gindreau, E. Garcia, R. Lopez, and J. Garcia. 2001. Complete nucleotide sequence of the genome of *Streptococcus pneumoniae* temperate bacteriophage MM1, p. 481. ASM 101st General Meeting. ASM Press, Washington, D.C.
59. Gill, D. M., T. Uchida, and R. A. Singer. 1972. Expression of diphtheria toxin genes carried by integrated and nonintegrated phage beta. Virology 50:664–668.
60. Goshorn, S. C., and P. M. Schlievert. 1989. Bacteriophage association of streptococcal pyrogenic exotoxin type C. J. Bacteriol. 171:3068–3073.
61. Groman, N. B., and M. Eaton. 1955. Genetic factors in *Corynebacterium diphtheriae* conversion. J. Bacteriol. 70:637–640.
62. Guan, S., D. A. Bastin, and N. K. Verma. 1999. Functional analysis of the O antigen glucosylation gene cluster of *Shigella flexneri* bacteriophage SfX. Microbiology 145:1263–1273.
63. Halperin, S. A., P. Ferrieri, E. D. Gray, E. L. Kaplan, and L. W. Wannamaker. 1987. Antibody response to bacteriophage hyaluronidase in acute glomerulonephritis after group A streptococcal infection. J. Infect. Dis. 155:253–261.
64. Hardt, W. D., L. M. Chen, K. E. Schuebel, X. R. Bustelo, and J. E. Galan. 1998. *S. typhimurium* encodes an activator of Rho GTPases that induces membrane ruffling and nuclear responses in host cells. Cell 93:815–826.
65. Hardt, W. D., H. Urlaub, and J. E. Galan. 1998. A substrate of the centisome 63 type III protein secretion system of *Salmonella typhimurium* is encoded by a cryptic bacteriophage. Proc. Natl. Acad. Sci. USA 95:2574–2579.
66. Hariharan, H., and W. R. Mitchell. 1976. Observations on bacteriophages of *Clostridium botulinum* type C isolates from different sources and the role of certain phages in toxigenicity. Appl. Environ. Microbiol. 32: 145–158.
67. Hayashi, T., T. Baba, H. Matsumoto, and Y. Terawaki. 1990. Phage-conversion of cytotoxin production in *Pseudomonas aeruginosa.* Mol. Microbiol. 4:1703–1709.
68. Hayashi, T., H. Matsumoto, M. Ohnishi, and Y. Terawaki. 1993. Molecular analysis of a cytotoxin-converting phage, phi CTX, of *Pseudomonas aeruginosa*: structure of the *attP-cos-ctx* region and integration into the serine tRNA gene. Mol. Microbiol. 7:657–667.
69. Head, S. C., M. Petric, S. Richardson, M. Roscoe, and M. A. Karmali. 1988. Purification and characterization of verocytotoxin 2. FEMS Microbiol. Lett. 51:211–216.
70. Holloway, B. W., and G. N. Cooper. 1962. Lysogenic conversion in *Pseudomonas aeruginosa.* J. Bacteriol. 84:1321–1324.
71. Holmes, R. K., and L. Barksdale. 1969. Genetic analysis of tox^{+} and tox^{-} bacteriophages of *Corynebacterium diphtheriae.* J. Virol. 3:586–598.
72. Huang, A., S. De Grandis, J. Friesen, M. Karmali, M. Petric, R. Congi, and J. L. Brunton. 1986. Cloning and expression of the genes specifying Shiga-like toxin production in *Escherichia coli* H19. J. Bacteriol. 166:375–379.
73. Hueck, C. J. 1998. Type III protein secretion systems in bacterial pathogens of animals and plants. Microbiol. Mol. Biol. Rev. 62:379–433.
74. Hull, A. E., D. W. K. Acheson, A. Donohue-Rolfe, G. T. Keusch, and P. Echeverria. 1991. Comparison of DNA probe and mitomycin C-enhanced immunoblot assay for the detection of Shiga-like toxin producing *E. coli* in stool specimens, p. 390. ASM 91st General Meeting. American Society for Microbiology, Washington, D.C.
75. Hull, A. E., D. W. K. Acheson, P. Echeverria, A. Donohue-Rolfe, and G. T. Keusch. 1993. Mitomycin immunoblot colony assay for detection of Shiga-like toxin-producing *Escherichia coli* in fecal samples: comparison with DNA probes. J. Clin. Microbiol. 31:1167–1172.
76. Hyder, S. L., and M. M. Streitfeld. 1978. Transfer of erythromycin resistance from clinically isolated lysogenic strains of *Streptococcus pyogenes* via their endogenous phage. J. Infect. Dis. 138:281–286.
77. Hynes, W. L., and J. J. Ferretti. 1989. Sequence analysis and expression in *Escherichia coli* of the hyaluronidase gene of *Streptococcus pyogenes* bacteriophage H4489A. Infect. Immun. 57:533–539.
78. Iida, T., A. Hattori, K. Tagomori, H. Nasu, R. Naim, and T. Honda. 2001. Filamentous phage associated with recent pandemic strains of *Vibrio parahemolyticus.* Emerg. Infect. Dis. 7:477–478.
79. Inoue, K., and H. Iida. 1970. Conversion of toxigenicity in *Clostridium botulinum* type C. Jpn. J. Microbiol. 14:87–89.
80. Inoue, K., and H. Iida. 1971. Phage-conversion of toxigenicity in *Clostridium botulinum* types C and D. Jpn. J. Med. Sci. Biol. 24:53–56.
81. Jackson, M. P., J. W. Newland, R. K. Holmes, and A. D. O'Brien. 1987. Nucleotide sequence analysis of the structural genes for Shiga-like toxin I encoded by bacteriophage 933J from *Escherichia coli.* Microb. Pathog. 2:147–153.
82. Jimenez-Lucho, V. E., and L. Leive. 1990. Role of the O-antigen of lipopolysaccharide in *Salmonella* in protection against complement action. *In* B. H. Iglewski, and V. L. Clark (eds.) The Bacteria, vol. XI: Molecular Basis of Bacterial Pathogenesis. Academic Press, San Diego.
83. Johns, M. B., and S. A. Khan. 1988. Staphylococcal enterotoxin B gene is associated with a discrete genetic element. J. Bacteriol. 170:4033–4039.
84. Johnson, E. A., and M. Bradshaw. 2001. *Clostridium botulinum* and its neurotoxins: a metabolic and cellular perspective. Toxicon 39:1703–1722.
85. Johnson, L. P., and P. M. Schlievert. 1984. Group A streptococcal phage T12 carries the structural gene for pyrogenic exotoxin type A. Mol. Gen. Genet. 194:52–56.

86. Kaneko, J., T. Kimura, Y. Kawakami, T. Tomita, and Y. Kamio. 1997. Panton-Valentine leukocidin genes in a phage-like particle isolated from mitomycin C-treated *Staphylococcus aureus* V8 (ATCC 49775). Biosci. Biotechnol. Biochem. 61:1960–1962.

87. Kaneko, J., K. Muramoto, and Y. Kamio. 1997. Gene of LukF-PV-like component of Panton-Valentine leukocidin in *Staphylococcus aureus* P83 is linked with *lukM.* Biosci. Biotechnol. Biochem. 61:541–544.

88. Karaolis, D. K. R., S. Somara, D. R. Maneval, J. A. Johnson, and J. B. Kaper. 1999. A bacteriophage encoding a pathogenicity island, a type-IV pilus and a phage receptor in cholera bacteria. Nature 399:375–379.

89. Karch, H., J. Heeseman, and R. Laufs. 1987. Phage-associated cytotoxin production by and enteroadhesiveness of enteropathogenic *Escherichia coli* isolated from infants with diarrhea in West Germany. J. Infect. Dis. 155:707–715.

90. Kehoe, M. A., V. Kapur, A. M. Whatmore, and J. M. Musser. 1996. Horizontal gene transfer among group A streptococci: implications for pathogenesis and epidemiology. Trends Microbiol. 4:436–443.

91. Kimmitt, P. T., C. R. Harwood, and M. R. Barer. 2000. Toxin gene expression by Shiga toxin-producing *Escherichia coli*: the role of antibiotics and the bacterial SOS response. Emerg. Infect. Dis. 6:458–465.

92. Kondo, I., and K. Fujise. 1977. Serotype B staphylococcal bacteriophage singly converting staphylokinase. Infect. Immun. 18:266–272.

93. Koonin, E. V. 1992. The second cholera toxin, Zot, and its plasmid-encoded and phage-encoded homologues constitute a group of putative ATPases with an altered purine NTP-binding motif. FEBS Lett. 312:3–6.

94. Kuzio, J., and A. M. Kropinski. 1983. O-antigen conversion in *Pseudomonas aeruginosa* PAO1 by bacteriophage D3. J. Bacteriol. 155:203–212.

95. Lacey, R. W. 1980. Evidence for two mechanisms of plasmid transfer in mixed cultures of *Staphylococcus aureus.* J. Gen. Microbiol. 119:423–435.

96. Laird, W., and N. B. Groman. 1976. Prophage map of converting corynebacteriophage beta. J. Virol. 19: 208–219.

97. Lanyi, B., and J. Lantos. 1976. Antigenic changes in *Pseudomonas aeruginosa* in vivo and after lysogenization in vitro. Acta Microbiol. Acad. Sci. Hung. 23:337–351.

98. Lazar, S., and M. K. Waldor. 1998. ToxR-independent expression of cholera toxin from the replicative form of CTXϕ. Infect. Immun. 66:394–397.

99. Liang-Takasaki, C. J., P. H. Makela, and L. Leive. 1982. Phagocytosis of bacteria by macrophages: changing the carbohydrate of lipopolysaccharide alters interaction with complement and macrophages. J. Immunol. 128: 1229–1235.

100. Lindsay, J. A., A. Ruzin, H. F. Ross, N. Kurepina, and R. P. Novick. 1998. The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in *Staphylococcus aureus.* Mol. Microbiol. 29:527–543.

101. Locke, A., and E. R. Main. 1931. The relation of copper and iron to production of toxin and enzyme action. J. Infect. Dis. 48:419–435.

102. Lui, P. V. 1969. Changes in somatic antigens of *Pseudomonas aeruginosa* induced by bacteriophages. J. Infect. Dis. 119:237–246.

103. Lyon, B. R., and R. Skurray. 1987. Antimicrobial resistance of *Staphylococcus aureus*: genetic basis. Microbiol. Rev. 51:88–134.

104. Madhubala, K. Prakash, and K. B. Sharma. 1981. Change in the 1O1 serotype of *Pseudomonas aeruginosa* after lysogenisation with bacteriophages. Indian J. Med. Res. 73:686–691.

105. Makino, K., K. Yokoyama, Y. Kubota, C. H. Yutsudo, S. Kimura, K. Kurokawa, K. Ishii, M. Hattori, I. Tatsuno, H. Abe, T. Iida, K. Yamamoto, M. Onishi, T. Hayashi, T. Yasunaga, T. Honda, C. Sasakawa, and H. Shinagawa. 1999. Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic *Escherichia coli* O157:H7 derived from the Sakai outbreak. Genes Genet. Syst. 74:227–239.

106. Marvaud, J. C., M. Gibert, K. Inoue, Y. Fujinaga, K. Oguma, and M. R. Popoff. 1998. botR/A is a positive regulator of botulinum neurotoxin and associated non-toxin protein genes in *Clostridium botulinum* A. Mol. Microbiol. 29:1009–1018.

107. Matsuda, M., and L. Barksdale. 1966. Phage-directed synthesis of diphtherial toxin in non-toxinogenic *Corynebacterium diphtheriae.* Nature 210:911–913.

108. Matsuda, M., and L. Barksdale. 1967. System for the investigation of the bacteriophage-directed synthesis of diphtherial toxin. J. Bacteriol. 93:722–730.

109. Matsuda, M., C. Kanei, and M. Yoneda. 1971. Temperature sensitive mutants of non-lysogenizing corynebacteriophage β^{hv}: their isolation, characterization and relation to toxinogenesis. Biken J. 14:119–130.

110. Mavris, M., P. A. Manning, and R. Morona. 1997. Mechanism of bacteriophage SfII-mediated serotype conversion in *Shigella flexneri.* Mol. Microbiol. 26:939–950.

111. McCormick, J. K., and P. M. Schlievert. 2000. Toxins and superantigens of group A streptococci, pp. 43–52. *In* V. A. e. a. Fischetti (ed.) Gram-Positive Pathogens. ASM Press, Washington, D.C.

112. McDonough, M. A., and J. R. Butterton. 1999. Spontaneous tandem amplification and deletion of the Shiga toxin operon in *Shigella dysenteriae* 1. Mol. Microbiol. 34:1058–1069.

113. McShan, W. M. 2000. The bacteriophages of group A streptococci, pp. 105–116. *In* V. A. e. a. Fischetti (ed.) Gram-Positive Pathogens. ASM Press, Washington, D.C.

114. Mirold, S., W. Rabsch, M. Rohde, S. Stender, H. Tschape, H. Russmann, E. Igwe, and W. D. Hardt. 1999. Isolation of a temperate bacteriophage encoding the type III effector protein SopE from an epidemic *Salmonella typhimurium* strain. Proc. Natl. Acad. Sci. USA 96: 9845–9850.

115. Miyamoto, H., W. Nakai, N. Yajima, A. Fujibayashi, T. Higuchi, K. Sato, and A. Matsushiro. 1999. Sequence analysis of Stx2-converting phage VT2-Sa shows a great divergence in early regulation and replication regions. DNA Res. 6:235–240.

116. Mizutani, S., N. Nakazono, and Y. Sugino. 1999. The so-called chromosomal verotoxin genes are actually carried by defective prophages. DNA Res. 6:141–143.
117. Montecucco, C., and G. Schiavo. 1995. Structure and function of tetanus and botulinum neurotoxins. Q. Rev. Biophys. 28:423–472.
118. Muschel, L. H., L. A. Ahl, and L. S. Baron. 1968. Effect of lysogeny on serum sensitivity. J. Bacteriol. 96:1912–1914.
119. Nakayama, K., S. Kanaya, M. Ohnishi, Y. Terawaki, and T. Hayashi. 1999. The complete nucleotide sequence of phi CTX, a cytotoxin-converting phage of *Pseudomonas aeruginosa*: implications for phage evolution and horizontal gene transfer via bacteriophages. Mol. Microbiol. 31:399–419.
120. Narita, S., J. Kaneko, J. Chiba, Y. Piemont, S. Jarraud, J. Etienne, and Y. Kanio. 2001. Phage conversion of Panton-Valentine leukocidin in *Staphylococcus aureus*: molecular analysis of a PVL-converting phage, ϕSLT. Gene 268:195–206.
121. Neely, M. N., and D. I. Friedman. 1998. Arrangement and functional identification of genes in the regulatory region of lambdoid phage H-19B, a carrier of a Shiga-like toxin. Gene 223:105–113.
122. Newland, J. W., and R. J. Neill. 1988. DNA probes for Shiga-like toxins I and II and for toxin-converting bacteriophages. J. Clin. Microbiol. 26:1292–1297.
123. Newland, J. W., N. A. Strockbine, S. F. Miller, A. D. O'Brien, and R. K. Holmes. 1985. Cloning of Shiga-like toxin structural genes from a toxin converting phage of *Escherichia coli.* Science 230:179–181.
124. Nizet, V., B. Beall, D. J. Bast, V. Datta, L. Kilburn, D. E. Low, and J. C. S. De Azavedo. 2000. Genetic locus for sreptolysin S production by group A streptococcus. Infect. Immun. 68:4245–4254.
125. Novick, R. P. 2000. Pathogenicity factors and their regulation, pp. 392–407. *In* V. A. e. a. Fischetti (ed.) Gram-Positive Pathogens. ASM Press, Washington, D.C.
126. Obrig, T. G., A. L. Baltch, T. P. Moran, S. P. Mudzinski, R. P. Smith, and F. Lutz. 1984. Effect of *Pseudomonas aeruginosa* cytotoxin on thymidine incorporation by murine splenocytes. Infect. Immun. 45:756–760.
127. Ostroff, S. M., P. I. Tarr, M. A. Neill, J. H. Lewis, N. Hargrett-Bean, and J. M. Kobayashi. 1989. Toxin genotypes and plasmid profiles as determinants of systemic sequelae in *Escherichia coli* O157:H7 infections. J. Infect. Dis. 160:994–998.
128. Pacheco, S. V., O. G. Gonzalez, and G. L. P. Contreras. 1997. The *lom* gene of bacteriophage λ is involved in *Escherichia coli* K12 adhesion to human buccal epithelial cells. FEMS Microbiol. Lett. 156:129–132.
129. Pappenheimer, A. M. 1977. Diphtheria toxin. Annu. Rev. Biochem. 46:69–94.
130. Pappenheimer, A. M., and S. J. Johnson. 1936. Studies in diphtheria toxin production. I. The effect of iron and copper. Br. J. Exp. Pathol. 17:335–341.
131. Pavia, A. T., C. R. Nichols, D. P. Green, R. V. Tauxe, S. Mottice, K. D. Greene, J. G. Wells, R. L. Siegler, E. D. Brewer, D. Hannon, et al. 1990. Hemolytic-uremic syndrome during an outbreak of *Escherichia coli* O157:H7 infections in institutions for mentally retarded persons: clinical and epidemiologic observations. J. Pediatr. 116: 544–551.
132. Plunkett, G., 3rd, D. J. Rose, T. J. Durfee, and F. R. Blattner. 1999. Sequence of Shiga toxin 2 phage 933W from *Escherichia coli* O157:H7: Shiga toxin as a phage late-gene product. J. Bacteriol. 181:1767–1778.
133. Projan, S. J. 2000. antibiotic resistance in the staphylococci, p. 463–470. *In* V. A. e. a. Fischetti (ed.), Gram-Positive Pathogens. ASM Press, Washington, D.C.
134. Ramirez, M., E. Severina, and A. Tomasz. 1999. A high incidence of prophage carriage among natural isolates of *Streptococcus pneumoniae.* J. Bacteriol. 181:3618–3625.
135. Reeve, J. N., and J. E. Shaw. 1979. Lambda encodes an outer membrane protein: the *lom* gene. Mol. Gen. Genet. 172:243–248.
136. Reeves, P. 1995. Role of O-antigen variation in the immune response. Trends Microbiol. 3:381–386.
137. Robbins, P. W., and T. Uchida. 1965. Chemical and macromolecular structure of O antigens from *Salmonella anatum* strains carrying mutants of bacteriophage ε^{34}. J. Biol. Chem. 240:375–383.
138. Robbins, P. W., and T. Uchida. 1962. Studies on the chemical basis of the phage conversion of O-antigens in the E-Group *Salmonellae.* Biochemistry 1:323–335.
139. Sako, T., S. Sawaki, T. Sakurai, S. Ito, Y. Yoshizawa, and I. Kondo. 1983. Cloning and expression of the staphylokinase gene of *Staphylococcus aureus* in *Escherichia coli.* Mol. Gen. Genet. 190:271–277.
140. Schmitt, M. P., and R. K. Holmes. 1991. Iron-dependent regulation of diphtheria toxin and siderophore expression by the cloned *Corynebacterium diphtheriae* repressor gene *dtxR* in *C. diphtheriae* C7 strains. Infect. Immun. 59:1899–1904.
141. Schmitt, M. P., E. M. Twiddy, and R. K. Holmes. 1992. Purification and characterization of the diphtheria toxin repressor. Proc. Natl. Acad. Sci. USA 89:7576–7580.
142. Shelton, C. B., D. R. Crosslin, J. L. Casey, S. Ng, L. M. Temple, and P. E. Orndorff .2000. Discovery, purification, and characterization of a temperate transducing bacteriophage for *Bordetella avium*. J. Bacteriol. 182: 6130–6136.
143. Shelton, C. B., L. M. Temple, and P. E. Orndorff. 2002. Use of bacteriophage Ba1 to identify properties associated with *Bordetella avium* virulence. Infect. Immun. 70:1219–1224.
144. Singer, R. A. 1973. Temperature-sensitive mutants of toxinogenic corynebacteriophage beta. I. Genet. Virol. 55:347–356.
145. Singer, R. A. 1973. Temperature-sensitive mutants of toxinogenic corynebacteriophage beta. II. Properties of mutant phages. Virology 55:357–362.
146. Skeeles, E. M., and L. H. Arp. 1997. Bordetellosis (turkey coryza). *In* B. W. e. a. Calnek (ed.) Disease of Poultry. Iowa State University Press, Ames, Iowa.
147. Skjold, S. A., W. R. Maxted, and L. W. Wannamaker. 1982. Transduction of the genetic determinant for streptolysin S in group A streptococci. Infect. Immun. 38:183–188.

148. Skorupski, K., and R. K. Taylor. 1997. Control of the ToxR virulence regulon in *Vibrio cholerae* by environmental stimuli. Mol. Microbiol. 25:1003–1009.
149. Spanier, J. G., and P. P. Cleary. 1980. Bacteriophage control of antiphagocytic determinants in group A streptococci. J. Exp. Med. 152:1393–1406.
150. Stanley, R. L., C. D. Ellermeier, and J. M. Slauch. 2000. Tissue-specific gene expression identifies a gene in the lysogenic phage Gifsy-1 that affects *Salmonella enterica* serovar typhimurium survival in Peyer's patches. J. Bacteriol. 182:4406–4413.
151. Stroeher, U. H., L. Bode, L. Beutin, and P. A. Manning. 1993. Characterization and sequence of a 33–kDa enterohemolysin (Ehly 1)-associated protein in *Escherichia coli.* Gene 132:89–94.
152. Temple, L. M. 2001.
153. Trucksis, M., J. E. Galen, J. Michalski, A. Fasano, and J. B. Kaper. 1993. Accessory cholera enterotoxin (Ace), the third toxin of a *Vibrio cholerae* virulence cassette. Proc. Natl. Acad. Sci. USA 90:5267–5271.
154. Ubukata, K., M. Konno, and R. Fujii. 1975. Transduction of drug resistance to tetracycline, chloramphenicol, macrolides, lincomycin and clindamycin with phages induced from *Streptococcus pyogenes*. J. Antibiot. (Tokyo) 28:681–688.
155. Uchida, T., D. M. Gill, and A. M. Pappenheimer. 1971. Mutation in the structural gene for diphtheria toxin carried by temperate phage β. Nat. New Biol. 233:8–11.
156. Unkmeir, A., and H. Schmidt. 2000. Structural analysis of phage-borne *stx* genes and their flanking sequences in Shiga toxin-producing *Escherichia coli* and *Shigella dysenteriae* type 1 strains. Infect. Immun. 68:4856–4864.
157. Vaca, S., J. Arce, G. Oliver, D. Arenas, and F. Arguello. 1989. FIZ15 bacteriophage increases the adhesion of *Pseudomonas aeruginosa* to human buccal epithelial cells. Rev. Lat. Am. Microbiol. 31:1–5.
158. van der Vijver, J. C., M. van Es-Boon, and M. F. Michel. 1972. Lysogenic conversion in *Staphylococcus aureus* to leucocidin production. J. Virol. 10:318–319.
159. van Horne, S. J., D. Bjornsen, P. Carpentier, and L. M. Temple. 2001. Investigation of expression of phage encoded pertussis toxin operon in *Bordetella avium* lysogens, abstract B–109, pp. 64–65. American Society for Microbiology, 101st General Meeting. ASM Press, Washington, D.C.
160. Van Wamel, W., A. Peshel, K. Van Kessel, and J. Van Strijp. 2001. CHIPS a chemotaxis inhibitory protein of *Staphylococcus aureus* is encoded by a prophage located pathogenicity island, abstract B–98, p. 62. American Society for Microbiology 101st General Meeting. ASM Press, Washington, D.C.
161. Veldkamp, K. E., H. C. Heezius, J. Verhoef, J. A. van Strijp, and K. P. van Kessel. 2000. Modulation of neutrophil chemokine receptors by *Staphylococcus aureus* supernate. Infect. Immun. 68:5908–5913.
162. Voelker, L. L., and K. Dybvig. 1999. Sequence analysis of the *Mycoplasma arthritidis* bacteriophage MAV1 genome identifies the putative virulence factor. Gene 233:101–107.
163. Voelker, L. L., K. E. Weaver, L. J. Ehle, and L. R. Washburn. 1995. Association of lysogenic bacteriophage MAV1 with virulence of *Mycoplasma arthritidis*. Infect. Immun. 63:4016–4023.
164. Wadolkowski, E. A., L. M. Sung, J. A. Burris, J. E. Samuel, and A. D. O'Brien. 1990. Acute renal tubular necrosis and death of mice orally infected with *Escherichia coli* strains that produce Shiga-like toxin type II. Infect. Immun. 58:3959–3965.
165. Wagner, P. L., D. W. K. Acheson, and M. K. Waldor. 2001. Human neutrophils and their products induce Shiga toxin production by enterohemorrhagic *Escherichia coli.* Infect. Immun. 69:1934–1937.
166. Wagner, P. L., D. W. K. Acheson, and M. K. Waldor. 1999. Isogenic lysogens of diverse Shiga toxin 2-encoding bacteriophages produce markedly different amounts of Shiga toxin. Infect. Immun. 67:6710–6714.
167. Wagner, P. L., J. Livny, M. N. Neely, D. W. K. Acheson, D. I. Friedman, and M. K. Waldor. 2002. Bacteriophage control of Shiga toxin 1 production and release by *Escherichia coli.* Mol. Microbiol. 44:957–970.
168. Wagner, P. L., M. N. Neely, X. Zhang, D. W. Acheson, M. K. Waldor, and D. I. Friedman. 2001. Role for a phage promoter in Shiga toxin 2 expression from a pathogenic *Escherichia coli* strain. J. Bacteriol. 183: 2081–2085.
169. Wagner, P. L., and M. K. Waldor. 2002. Bacteriophage control of bacterial virulence. Infect. Immun. 70: 3985–3993.
170. Waldor, M. K., and J. J. Mekalanos. 1996. Lysogenic conversion by a filamentous phage encoding cholera toxin. Science 272:1910–1914.
171. Waldor, M. K., and J. J. Mekalanos. 1996. *Vibrio cholerae*: molecular pathogenesis, immune response, and vaccine development. *In* L. J. Paradise, et al. (eds.) Enteric Infections and Immunity. Plenum Press, New York.
172. Weeks, C. R., and J. J. Ferretti. 1984. The gene for type A streptococcal exotoxin (erythrogenic toxin) is located in bacteriophage T12. Infect. Immun. 46:531–536.
173. Willshaw, G. A., H. R. Smith, S. M. Scotland, and B. Rowe. 1985. Cloning of genes determining the production of vero cytotoxin by *Escherichia coli.* J. Gen. Microbiol. 131:3047–3053.
174. Winkler, K. C., J. de Waart, and C. Grootsen. 1965. Lysogenic conversion of staphylococci to loss of beta-toxin. J. Gen. Microbiol. 39:321–333.
175. Wong, C. S., S. Jelacic, R. L. Habeeb, S. L. Watkins, and P. I. Tarr. 2000. The risk of the hemolytic-uremic syndrome after antibiotic treatment of *Escherichia coli* O157:H7 infections. N. Engl. J. Med. 342:1930–1936.
176. Wood, M. W., R. Rosqvist, P. B. Mullan, M. H. Edwards, and E. E. Galyov. 1996. SopE, a secreted protein of *Salmonella dublin*, is translocated into the target eukaryotic cell via a *sip*-dependent mechanism and promotes bacterial entry. Mol. Microbiol. 22:327–338.

177. Wright, A. 1971. Mechanism of conversion of the *Salmonella* O antigen by bacteriophage ε^{34}. J. Bacteriol. 105:927–936.
178. Xiong, G., J. Lin, P. Oepen, D. Senerwa, and F. Lutz. 1998. Control of effective expression of the phage phiCTX-encoded *ctx* gene in *Pseudomonas aeruginosa* by a promoter upstream of the *cos* site. Mol. Gen. Genet. 257:249–254.
179. Yamaguchi, T., T. Hayashi, H. Takami, K. Nakasone, M. Ohnishi, K. Nakayama, S. Yamada, H. Komatsuzawa, and M. Sugai. 2000. Phage conversion of exfoliative toxin A production in *Staphylococcus aureus*. Mol. Microbiol. 38:694–705.
180. Yokoyama, K., K. Makino, Y. Kubota, M. Watanabe, S. Kimura, C. H. Yutsudo, K. Kurokawa, K. Ishii, M. Hattori, I. Tatsuno, H. Abe, M. Yoh, T. Iida, M. Ohnishi, T. Hayashi, T. Yasunaga, T. Honda, C. Sasakawa, and H. Shinagawa. 2000. Complete nucleotide sequence of the prophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic *Escherichia coli* O157:H7 strain derived from the Sakai outbreak. Gene 258:127–139.
181. Zabriskie, J. B. 1964. The role of temperate bacteriophage in the production of erythrogenic toxin by group A streptococci. J. Exp. Med. 119:761–779.

48

Phage Therapy

CARL R. MERRIL
DEAN SCHOLL
SANKAR ADHYA

The ability of bacteriophage (phage) to replicate exponentially and lyse pathogenic strains of bacteria suggests that they should play a vital role in our armamentarium for the treatment of infectious diseases. However, in spite of an initial enthusiasm, early clinical applications resulted in a negative shift of opinion concerning the therapeutic potential of phage. There are a number of factors that may have been responsible for this rejection of the use of phage as antibacterial therapeutic agents, particularly in countries that require certification based on the results of efficacy and pharmacokinetic studies in animals and humans. These factors include an initial lack of understanding of the relatively narrow host range of phage and an inability to purify phage preparations from bacteria products and debris. These contaminating materials often include bacterial exo- and endotoxins along with bacterial cellular components that tend to inactivate phage preparations when they are stored without further purification (60). Another major factor that affected the development of phage therapy was the successful introduction of antibiotics effective against a broad range of bacterial strains. With such antibiotics physicians could often successfully treat infections even before they determined the causative bacterial strain. The narrow host range of phage made duplication of such a practice questionable at best.

Despite the current widespread use of antibiotics, an ever-increasing prevalence of antibiotic-resistant bacteria-suggests that phage therapy merits reconsideration. In addition, knowledge gained since the initial phage therapeutic applications, concerning phage genetics, physiology, and molecular biology, should provide beneficial information for current efforts to develop phage into a reliable therapeutic agent. However, there are still a number of gaps in our knowledge, including information on the interaction of phage with mammalian organisms including the reactions of innate and active mammalian immune systems to phage. We also need to develop methods to understand and improve on the pharmacokinetic behavior of phage. In addition, we need techniques to facilitate the rapid determination of the best phage strain to use for each specific clinical infection before we can make full use of the therapeutic potential of phage.

As the discovery of antibiotics was one of the major events that overshadowed the development of phage as an antibacterial therapeutic agent, we have divided the following historical introduction into pre-antibiotic and antibiotic eras. See chapter 1 for a discussion of phage history from the perspective of the development of the discipline of molecular biology.

Pre-antibiotic Era Phage Therapy

Shortly after phage was discovered by Twort in 1915 (77), Felix d'Hérelle championed the concept of using them to treat bacterial infections (19). d'Hérelle's first efforts were concentrated on the treatment of avian typhosis in chickens and shigella dysentery in rabbits. Following his reported success in these applications of phage as antibacterial therapeutic agents, he extended the use of phage to the treatment of bacillary dysentery in human infections. In pursuing his phage therapy studies he traveled around the world, stimulating both basic and clinical phage research (75). While many of d'Hérelle's ideas concerning phage have proven correct, one idea that he proposed may have resulted in some of the clinical failures of phage therapy. d'Hérelle suggested that there might be only one bacteriophage that could adapt to many bacterial hosts. He even named this perceived phage capacity, the "unicity of the bacteriophage" (80). This belief that there was but one phage that could adapt to all bacterial strains may have led clinicians at that time to use inappropriate phage strains. In 1928, d'Hérelle was appointed to a professorship in the bacteriology department at Yale University School of Medicine (80). During his tenure at Yale he conducted a number of phage studies including efforts to "adapt" *Staphylococcus* phage to resist inhibition by factors in human serum (20). Following his resignation from his position at Yale in 1934, d'Hérelle played a major role in the establishment of an institute in Tbilisi, in Soviet Georgia. This Institute produced large quantities of phage for antibacterial therapy during and immediately following the Second World War and today is still actively pursuing phage therapy applications.

Phage were also used as antibacterial therapeutic agents in Poland and France, and they were distributed for clinical practice in the United States by a number of major pharmaceutical companies. In 1932, one European laboratory was reported to be distributing 50 liters of phage a day (39). Phage preparations were marketed by three major US Pharmaceutical companies. Eli Lilly & Co. sold "Staphylo jel," and other phage "jel"-labeled products, for streptococcus and colon bacillus. Phage products, marketed by E. R. Squibb & Sons and the Swan-Myers division of Abbott Laboratories included a bacteriophage filtrate preparation for staphylococcus and a combined bacteriophage filtrate preparation for staphylococcus and colon bacillus, respectively (74). Problems with some of these commercial phage preparations were found to be caused by the use of organomercury compounds as preservatives. Such preservatives often resulted in loss of phage activity during storage. Variations in the phage strains from one lot to another but marketed under the same label served to further undermine the confidence of clinicians who might otherwise have used these preparations (74). Additional problems in early therapeutic applications of phage may have occurred because of the lack of refrigeration or adequate phage purification. Most phage preparations intended for therapeutic applications were purified only by passing the lysate through filters fine enough to remove the host bacteria. While such purification reduces the risk of bacterial infections, it does not remove bacterial debris that can include bacterial endo- and exotoxins, and these contaminants can result patient morbidity and mortality. Phage interactions with this material during storage of phage preparations can also result in loss of activity. These were some of the issues that resulted in the establishment of a Council on Pharmacy and Chemistry by the American Medical Association. This Council concluded that the phage therapy was plagued by a lack of basic understanding and standards for purity or effectiveness (23). A recent review by Ho (32) presents additional information pertaining to this period.

Researchers interested in phage therapy were also concerned that even when phage with demonstrated in vitro antibacterial effects were used for clinical infections, factors present in the serum, tissue debris, cellular components, etc., would inhibit bacterial lysis by phage. They suggested that given these issues any positive effect of phage therapy on the course of an infection was probably due to stimulation of specific antibacterial immunity, and/or nonspecific phagocytic activity (22, 24).

By 1937, the state of affairs had deteriorated to the point that researchers such as Asheshov and his colleagues stated that "no satisfactory evidence has yet been obtained that a phage exerts any significant effect on the course of an experimental infection" (5). Despite their skepticism, Asheshov and his colleagues recognized that part of the problem was associated with the difficulty of repeating experimental results due to failures of experimental design. By using a specific strain of bacterium and phage strains that were shown to be active in vitro against the selected bacterial strain, these researchers demonstrated the ability of the phage to rescue mice injected intraperitoneally with lethal concentrations of "*Bacter. Typhosum*." They were able to rescue a significant number of mice even when the phage injections were given as late as 4 hours after the bacterial injection. Without phage treatment most of the animals died within 24 hours. Furthermore, they demonstrated that heat-killed phage and nonspecific phage strains were not able to rescue the animals. These experiments clearly showed that phage that display antibacterial activity against a particular bacterial strain in vivo may serve as an antibacterial therapeutic agent for the treatment of an animal with a systemic infection with that bacterial strain. They also demonstrated that the antibacterial effect of the phage is due to the physiological functions of the phage since heat-killed and nonspecific phage could effect no such rescue of infected animals.

Dubos et al. (22), at Yale University, addressed the concerns that factors in blood, tissue, and bile might interfere with the lytic activity of phage and that such interference would render phage impotent as antibacterial therapeutic agents. They demonstrated that such interference effects, if present, were minimal as they were able to rescue mice infected intracerebrally with *Shigella dysenteriae* by injecting anti-Shiga phage into the general circulation. In these experiments they also observed a correlation between an increase in phage titer observed in the blood of infected animals and their rescue, suggesting that the rescue of the animals was due to phage functions, including replication, and that interfering factors, if present, were insufficient to inhibit the beneficial effects of phage as a potential antibacterial therapeutic agents (see chapter 5 for a discussion of phage population growth and its impact on bacterial population density).

These positive developments came too late to generate much enthusiasm for phage therapy in the Western world. By this time antibiotics were proving superior because of their activity against a broad range of bacterial hosts and in most cases their robust storage characteristics. While the use of phage therapy waned in Western medical practice it continued to be employed in Eastern Europe and parts of Asia. This was due in part to the restriction of information concerning the development of antibiotics, particularly penicillin, by the British and American governments in the early phase of the Second World War. The Soviet and Polish phage literature concerning the therapeutic use of phage has been extensively reviewed (1). The reviewers note that the Soviet and Polish medical researchers studied the efficacy of phage therapy almost exclusively by qualitative clinical assessments of patients. Details of phage dosages and clinical criteria were reported in a "sketchy" manner. For these reasons, most of the studies from Eastern Europe will not meet current standards in countries that require certification based on the results of efficacy and pharmacokinetic studies in animals and humans.

Antibiotic Era Phage Therapy

It was 30 years after the Dubos' work before animal studies were again performed to investigate the efficacy of phage to treat bacterial infections. This hiatus was due in large part to the success of antibacterial chemotherapeutics such as the sulfonamides, discovered in 1935, followed by the antibiotics during and following the Second World War. However, some of the resistance may have been due in part to the effect of theoretical explanations that arose to explain the perceived failure of phage therapy. These are best captured by the following statement from Stent's 1963 book, *Molecular Biology of Bacterial Viruses* (73):

> Just why bacteriophages, so virulent in their antibacterial action in vitro, proved to be so impotent in vivo has never been adequately explained. Possibly the immediate antibody response of the patient against the phage protein upon hypodermic injection, the sensitivity of the phage to inactivation by gastric juices upon oral administration, and the facility with which (as we shall see presently) bacteria acquire immunity or sport resistance against phages, all militated against the success of phage therapy.

As suggested by Stent, antibodies, produced by the adaptive immune system, may be of importance for the inactivation of phage, particularly in individuals repeatedly exposed to a specific phage strain. This ability of phage to provoke an antibody response in normal individuals has been used over the past three decades by Ochs and his colleagues, who employ the phage ϕX174, to study normal individuals and patients with immune deficiencies (56, 57). They demonstrated that the adaptive immune system of normal individuals who are naive to a particular phage strain requires a few days to develop a detectable antibody level and about 2 weeks for a maximal antibody response. However, the innate immune system has been shown to be able to rapidly eliminate phage administered systemically. This was investigated in an experiment, published in 1973, that demonstrated that phage injected systemically in germ-free mice were removed rapidly, by the liver and spleen (reticuloendothelial system (RES) of the innate immune system), from the circulatory system even though these mice displayed no antibody activity against the phage (30). The authors suggested that this rapid elimination of phage in intact animals "may explain the limited success reported for the phage treatment of infectious diseases." They also suggested that the rapid rate of phage elimination could be slowed by overwhelming the RES with colloidal particles.

Recently a less intrusive method was discovered to circumvent this rapid systemic elimination of phage. This method employs genetic selection to find mutant phage strains with reduced rates of clearance by the RES by employing serial passage techniques for the selection of such variants (50). Infected mice treated with these long-circulating phage variants recovered more rapidly and their symptoms were less severe than those mice treated with wild-type phage. In these experiments, long-circulating lytic-mutant λ and P22 phage (reviewed in chapters 27 and 29, respectively) were developed to treat mice infected intraperitoneally with *Escherichia coli* and *Salmonella typhimurium*, respectively. In the case of λ phage, the mutants displayed a single base change in the capsid *E* gene that resulted in the substitution of a lysine for the normal glutamic acid residue in this capsid protein. Given the large number of copies of this protein in the phage capsid, the resulting alkaline shift associated with the mutant phage may have been associated with the capacity of this phage to remain in the circulatory system for an extended period.

Another concern, that "bacteria acquire immunity or sport resistance against phage," as expressed by Stent, was addressed experimentally by Smith and Huggins (67). As predicted by Stent, Smith and Huggins found phage-resistant bacterial mutants following the use of a single dose of an *E. coli* K1 strain-specific phage to treat mice that had been infected, either intramuscularly or intracerebrally, with a K1 strain of *E. coli*. Interestingly, the phage treatment of these infected mice still proved to be more efficacious than treatment with antibiotics. In fact, phage-resistant bacterial mutations occurred with less frequency following phage therapy than the antibiotic-resistant mutants that appeared after antibiotic therapy. Furthermore, the phage-resistant bacterial mutants that were observed were found to be lacking the K1 capsule (the receptor for phage attachment) which is associated with a loss of pathogenicity.

Smith and Huggins' experiments, demonstrating that phage injected intramuscularly could be used to treat an intracerebral infection, corroborated the findings of Dubos et al. (22) obtained 39 years earlier, in which phage injected into the circulatory system were successfully used to treat an intracerebral *Shigella dysenteriae* infection in mice.

These observations may help to mitigate concerns that phage, due to their larger size, will not be as effective in treating tissue infections as the lower molecular weight antibiotics (42).

Experiments by Soothill (71) showed that the lytic development of the phage is a critical parameter in the ability of phage to control bacteria in vivo. In separate experiments, animals were given intraperitoneal injections of either *Acinetobacter baumanii* or *Pseudomonas aeruginosa* and then treated successfully using phage specific for each. However, in a similar experiment with *Staphylococcus aureus*, treatment with phage failed. *Staphylococcus* phage used were significantly less active in in vitro than the phages for the other organisms.

Experimentally induced septicemia and meningitis in chickens and calves have also been successfully treated with phage (7). In these studies, *E. coli* K1 strains were used along with the K1-specific R phage (the same phage strain as that used by Smith and Huggins). The phage was able to rescue chickens even when the administration was delayed until the onset of symptoms. As in the work by Dubos et al., 55 years earlier, in vivo phage multiplication was found in the brains of infected animals, even when the phage were injected intramuscularly. It is not clear whether the capacity of the phage to cross the blood–brain barrier was due to effects from an inflammatory response to the bacterial infection or whether phage can normally cross this barrier. It was also found that phage taken up by the spleen persisted in significant numbers for several days after injection, corroborating the findings of Geier et al. (30).

Much of the interest in reviving phage therapy has been fuelled by the appearance of antibiotic-resistant bacteria. In this regard, Biswas et al. (10) showed that phage specific for vancomycin-resistant *Enterococcus faecium* could rescue mice that were infected intraperitoneally with bacteria. If titers of phage equivalent to the titers of infecting bacteria were given 45 minutes after infection, 100% of the mice were rescued, and even when treatment was delayed for 24 hours, when the mice were moribund, 50% could still be rescued. Phage may also be used to treat antibiotic-resistant intracellular pathogens. Broxmeyer et al. (12) have demonstrated that it is possible to use *Mycobacterium smegmatis*, an avirulent mycobacterium, as a vector to deliver the lytic phage TM4 to treat intracellular mycobacterium infections (with either *Mycobacterium avium* or *Mycobacterium tuberculosis*) in macrophages. These examples of efforts to develop phage therapy represent cases where specific needs exist for the treatment of clinical infections that are not treatable by current methods. As the occurrence of drug-resistant pathogens increases, we expect to see an increase in efforts to develop phage into a viable alternative to antibiotics.

All the work described above involved the treatment of systemic infections using phages injected either intramusculary or intraperitoneally. Phage have been demonstrated to be effective therapeutic agents for the treatment of nonsystemic infections. Several studies have shown that gastrointestinal infections can be treated by oral administration of phage. Smith and Huggins (67) and Smith et al. (68, 69) showed that various phages could protect calves, pigs, and lambs from gastrointestinal infections of enteropathogenic strains of *E. coli*. In addition, Ramesh et al. (59) found that oral administration of a bacteriophage specific for *Clostridium difficile* could prevent ileocecitis caused by this organism in hamsters. Since *C. difficile* colonization of the gut is a common consequence of extended antibiotic treatment due to destruction of the normal gut flora, phage used in such a manner may prove to be useful in conjunction with antibiotic therapy.

A study by Soothill (72) showed that bacteriophage could prevent destruction of skin grafts by *Pseudomonas aeruginosa.* This bacterium commonly colonizes burn wounds and is often difficult to treat with antibiotics. There is an immediate need for a treatment of this problem, and topical application of phages to burn wounds with *Pseudomonas* infections may be an attractive way to reintroduce phages into the modern clinical setting.

Terrestrial animals are not the only candidates for phage therapy. Recent studies have shown that phage can be used to treat bacterial diseases of fish in aquaculture (53). In addition, phages have recently been shown to be effective in the treatment of bacterial blight of geranium (26) and bacterial spot on tomatoes (25).

Selection and Characterization of Therapeutic Phage Factors

Phage are the most abundant entities on the planet (there are estimated to be more than 10^{30} phage particles (13); see, for example, chapter 33 for discussion of this abundance). However, only a few phage strains will prove to be effective as therapeutic antibacterial agents. There are a number of factors that can affect the therapeutic efficacy of phage chosen for use as antibacterial therapeutic agents. Studies of phage multiplication in bacteria in vitro may provide information as to bacterial host range and a good first approximation of whether the phage may be appropriate for a particular clinical infection. However, observation of phage multiplication in defined culture media does not take into account the interactions of the phage with the bacteria in the clinical environment. Bacterial gene expression and phenotype may be affected by numerous variables in the clinical milieu, ranging from differences in the basic nutrients to altered physiological parameters including pH and ionic strength. In addition, clinical infections have the added complication of interactions between the innate and active immune systems with both the infectious bacteria and the therapeutic phage.

While phage therapy, as described above, has had a relatively long, but checkered, history, careful scientific study of factors associated with therapeutic efficacy has yet to be carried out.

Phage chosen as antibacterial therapeutic agents need to be well characterized, the genome sequenced, and much of the biology of the phage well understood before it is developed for therapeutic use. Phage host range, virulence, stability, interaction with both innate and active immune systems, as well as the phage's possible capacity to lysogenize and transduce, must be understood to provide the greatest chance of efficacy and safety.

In addition, therapeutic phage strains need to be tested and selected for their ability to function in the milieu of the human physiological systems. Interactions with both the innate and active immune system need to be minimized for phage strains that are employed for the treatment of systemic infections and pharmacokinetic parameters need to be determined.

Host Specificity of Phage

The host range of phage is generally narrower than that found in antibiotics selected for clinical applications. Most phages are specific for one species of bacteria and many are only able to lyse specific strains within a species (nevertheless, see chapter 46). The limited host range of phage can be both an advantage and a disadvantage in phage therapy.

The advantage of a narrow phage host range is that the use of such phage in antibacterial therapy results in less harm to the normal body flora and ecology. In contrast, antibiotics with their ability to affect a wide range of bacterial strains often disrupt the normal gastrointestinal flora. Such side effects of antibiotic therapy can result in opportunistic secondary infections by organisms such as *Clostridium difficile* (8). While this type of side effect should not be a problem with phage therapy, the narrow phage host range does require a means of determining the specific phage strain needed for each infection that is to be treated. This requirement for the use of phage as an antibacterial agent presents two major problems in the current clinical setting. The first problem is the need to have a battery of well-characterized phages available for a broad range of pathogens; second, there must be a timely method available to determine which phage strain will be effective for a given infection (as discussed below under "Methods for the Rapid Determination of Phage Specific for Infecting Pathogen").

Many different phage strains will need to be identified, characterized, and developed to cover even a proportion of the bacterial diseases that could be good candidates for phage therapy. However, using current molecular techniques it may be possible to enhance the host range of some phage, thus reducing the number of phages that need to be developed. For example, it has been found that coliphage K1-5 is a "dual" specificity phage that encodes two different tail proteins allowing it to attack and replicate on both K1 and K5 strains of *E. coli* (62). One tail protein found on phage K1-5 is a lyase protein, similar to that of phage K5 (specific for the K5 polysaccharide capsule) and a second tail protein found on this phage is an endosialidase similar to a tail protein found in phage K1E (specific for the K1 polysaccharide capsule). In addition, the genomic region encoding these proteins is almost identical to the genomic construct found in the *Salmonella* phage SP6 which codes for a protein that binds to the *Salmonella* O-antigen (figure 48-1) (63).

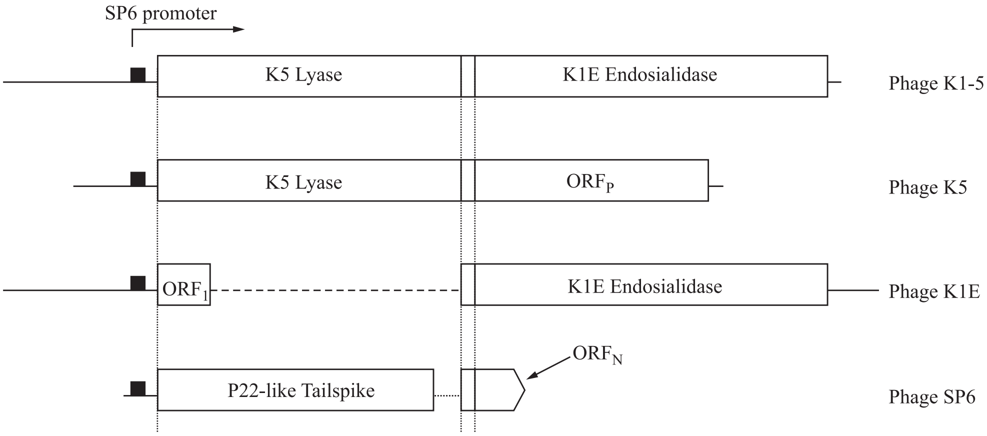


Figure 48-1 Phage tail protein genes. Diagram of tail protein genome-encoding regions for the coliphages K1, K5, K1-5, and the *Salmonella* phage SP6. All these phages share a similar promoter region and an intergenic region with a putative transcription terminator. This "modular" genomic construct suggests that a horizontal gene transfer mechanism for host range variation in nature that can be adapted for phage to be used as therapeutic antibacterial agents. These phages display additional qualities necessary for phages that will be used in antibacterial therapy: they produce progeny phage with a large burst size, and also show little if any loss of titer on storage.

The observation of a similar tail genome motif in both the *Salmonella* phage SP6 and the coliphage K1E, K5, and K1-5 suggests that this genomic construct might serve in the development of a modular phage platform that could operate over a wide bacterial host range. The development and use of such a "modular" phage would also save considerable time and effort over that required for characterizing completely new phages for each bacterial strain.

Other mechanisms have been found in nature to expand the bacterial host range of phage. These include the site-specific recombination systems that permit phage to switch between alternative tail fiber proteins (61) and the reverse transcriptase by which a *Bordetella* phage provides for variations in its tail fiber proteins (44). It may be possible to adapt these mechanisms to extend the host range of therapeutic phages.

Parameters besides tail fibers and their interaction with cellular receptors can be important for host specificity. Restriction/modification systems may limit the host range of phage in some bacterial strains. It may be possible to address this problem by engineering phages with genomes that do not contain restriction sites recognized by the nonpermissive host. Alternatively, phages could be produced in bacterial strains that provide DNA modification(s) that allow the phage to escape restriction in the targeted bacteria strain. Another approach that might be employed to address the restriction/modification problem when engineering a phage is exemplified by a mechanism used by the phage T7. This phage expresses a gene, *0.3*, early in the infection process that codes for a protein which is a potent inhibitor of type I DNA restriction and modification enzymes (6, 51). A construct containing this gene might be adapted for use in other phage strains or it may be possible to modify T7 to expand its bacterial host range for *E. coli* infections (phage T7 biology is reviewed in chapter 20).

In some cases, phage may fail to replicate in a particular host because they lack one or two genes essential for the replication of the phage. Such gene(s) can be identified and then incorporated in the phage genome. For example, phage λ does not normally replicate in *Salmonella*. However, when a λ phage library containing copies of the *E. coli* genome were tested, it was found that λ phage carrying the *E .coli nusA* gene could replicate in a *Salmonella* strain, provided that the receptor protein for λ attachment is already expressed in the *Salmonella* strain (S. Adhya, *unpublished observations*).

In addition to the factors addressed above, bacteria grown with standard laboratory protocols may not behave the same when they are growing in the milieu of an infection. This point was emphasized in a review by Hollon (33) in which he cited observations of Karakawa that *Staphylococcus aureus* rarely expresses the capsular polysaccharides found in clinical isolates when the bacteria are grown in the laboratory. Given such a change in the bacterial capsule, a phage discovered using bacteria grown in the laboratory may not be able to multiply in the same bacterial strain in an infected animal. In the early phage literature there are reports of body fluids (serum, pus, ascitic fluids, cerebrospinal fluid, urine, and bile) inhibiting the infectivity of phage against typhoid, colon bacilli, and staphylococci (14, 17). More recently, it has been reported that phage infecting certain strains of *E. coli* that are not expressing the cell surface protein, Ag43, in standard laboratory growth media, may be inhibited by concentrations of bile salts similar to those found in the gastrointestinal tract (28). The Ag43 protein is a phase-variable protein whose expression is associated with *E. coli* biofilm formation (18). Recognition of these problems is important in isolating phage for clinical applications.

Undesirable Phage Genes

While phage can be used to treat bacterial infections they can also play a major role in bacterial pathogenesis. A number of phage genes have been discovered that encode toxins, or factors that enhance bacterial virulence. They may also contribute, through transduction, to the transmission of antibiotic resistance genes (81). It may be possible to reduce the occurrence of such adverse effects by sequencing the genome of phages of interest for therapeutic applications and using this sequence information to search for homologies with known toxin genes, islands of pathogenicity or genes that foster integration of DNA into the bacterial genome. Known phage-encoded toxin genes are summarized in table 48-1; see also chapter 47 for a broader discussion of the role of phage in bacterial pathogenicity.

The presence of such toxin genes or genes with similar sequences can be found be found by searching phage genomes against GenBank online using the Basic Local Alignment Search Tool, BLAST (2). In addition, this approach can be used to search for drug-resistance genes, phage genomic-integration factors, or other potential genes that may increase the virulence of a bacterial strain. Such BLAST searches take into account only similarities to known genes and it can certainly be assumed that there are other, as yet unidentified toxins and potentially detrimental genes that do not have sequence similarity to anything currently in the databases. However, knowledge of toxins, drug resistance, and other potentially troublesome genes is increasing rapidly as is the number of completely sequenced phage and bacterial genomes. For these reasons, such database searches will become increasingly useful and they should help to assure that phage chosen for use as antibacterial therapeutic agents are free of genes that might potentially damage bacterially infected humans, animals, or plants being treated with phage therapy.

Table 48-1 Phages that Carry Toxin Genes (adapted from 11)

Bacteria	Phage	Gene	Gene product/phenotype
Escherichia coli O157:H7	933,H-19B	*stx*	Shiga toxins
	ϕFC3208	*hly2*	Enterohemolysin
	λ	*lom*	Serum resistance
	λ	*bor*	Host cell envelope protein
Shigella flexneri	Sfi6	*oac*	O-antigen acetylase
	sfII,sfV,sfX	*gtrII*	Glucosyl transferase
Salmonella enterica	SopEϕ	*sopE*	Type III effector
	Gifsy-2	*sodC-1*	Superoxide dismutase
	Gifsy-2	*nanH*	Neuraminidase
	Gifsy-1	*gipA*	Insertion element
	ε^{34}	*rfb*	Glucosylation
Vibrio cholerae	CTXϕ	*ctxAB*	Cholera toxin
	K139	*glo*	G-protein like
	VPIϕ	*tcp*	TCP pilin
Pseudomonas aeruginosa	CTXϕ	*ctx*	Cytotoxin
Clostridium botulinum	C1	*C1*	Neurotoxin
Staphylococcus aureus	NA	*see,sel*	Enterotoxin
	ϕ13	*entA,sak*	EnterotoxinA, staphylokinase
	TSST-1	*tst*	Toxic shock syndrome-1
Streptococcus pyogenes	T12	*speA*	Erythrogenic toxin
Corynebacterium diphtheriae	β-phage	*tox*	Diphtheria toxin

Pharmacokinetics of Phage Therapy

Phage in Mammalian Host

Pharmacokinetic data concerning phage therapy are still in a rudimentary state despite the long history of phage use and study. Many early clinical applications of phage therapy employed oral administration of phage preparations with little or no effort to determine phage uptake or distribution. While oral administration may have diminished the possible side effects from contaminants, including endo- and exotoxins that are often present in filter-"purified" phage preparations, it may not have been the most effective route for the treatment of systemic infections. Determination of the most effective therapeutic regime(s) for phage therapy requires pharmacokinetic information.

There were some early practitioners of phage therapy who recognized the need to learn the fate of phage injected into animals. However, these early researchers generally employed qualitative methods and they only reported whether lysis had occurred following the incubation of ground-up tissue or drops of blood with the host bacteria in liquid media. Despite these limitations, such efforts led to the observation in 1921 that phage injected into the circulatory system of rabbits could still be found in the spleen long after no trace of phage could be found in other organs or in the blood (3). This finding was corroborated 12 years later, in 1933, in an experiment in which, 3 days after the intravenous injection of a rabbit with phage, the animal was killed and the liver, spleen, and blood examined for phage. The liver and spleen, "crushed to a pulp in a mortar," and a sample of blood were independently incubated in growth media with the host bacteria. At that time phage could no longer be detected in the blood or the liver but it could be found in the spleen (24). One of the first quantitative studies of the fate of phage in animals was performed by Nungester and Watrous (55). They reported that following the intravenous inoculation of 10^9 plaque forming units (PFU) of a *Staphylococcus* phage into albino rats, the titer in the blood dropped to 10^5 PFU in 5 minutes and about 4×10^1 PFU in 2 hours. This rapid elimination of phage from the circulatory system was attributed to the organs of the RES primarily the liver and spleen.

In experiments using T4 phage as a probe of the immune system, Inchley (35) found that the liver phagocytosed more than 99% of the phage within 30 minutes after inoculation and that it removed 12 times as much phage as the spleen, as measured by the uptake of ^{51}Cr-labeled phage. Additional studies demonstrated that the liver inactivated the phage at a higher rate than the spleen, as measured by PFUs of phage that could be detected in these organs and the rate of loss of ^{131}I-labeled T4 phage in these organs (T4 biology is reviewed in chapter 18).

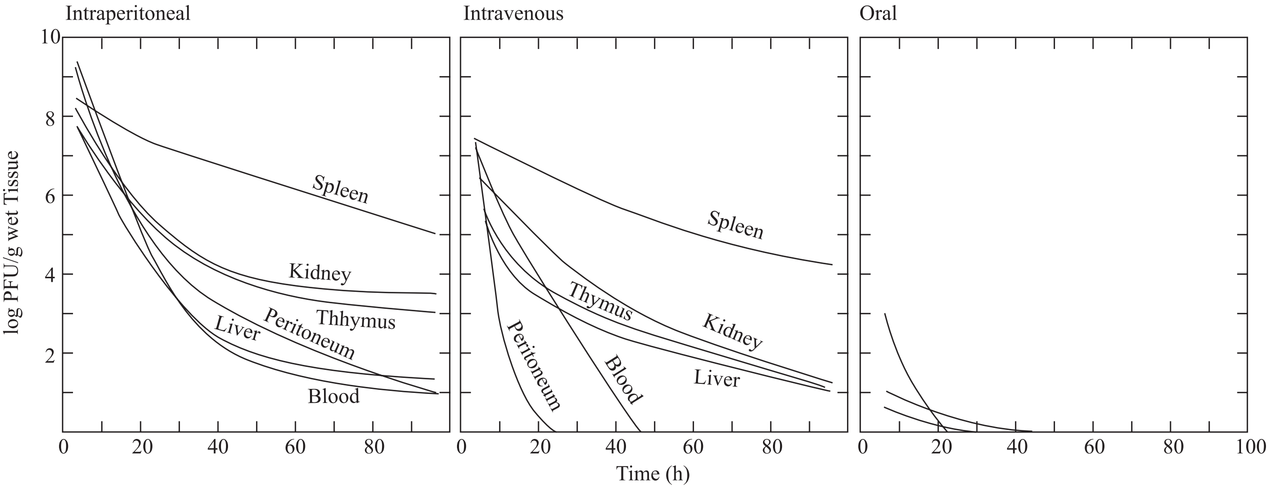


Figure 48-2 The distribution of phage plaque forming units in mice following various routes of administration. This graph was adapted from data from a 1973 experiment in which germ-free mice were inoculated with a single dose of 2 × 10^{12} plaque forming units (PFU) of λ phage. In these experiments oral administration of phage resulted in the detection of a systemic level of phage tissue titers that were 7 to 8 orders of magnitude lower than that achieved by systemic administration of phage (30).

A study of the distribution of phage was reported in the previously mentioned study by Geier et al. (30) that employed germ-free mice. These mice had no detectable antibody activity to the λ phage. Despite the lack of these antibodies the animals also displayed rapid elimination of the phage from the circulatory system and retention of active phage, as measured by PFU, in the spleen (figure 48-2). As there were no detectable antibody levels for the phage employed in these experiments, this initial reaction to phage by the animals must be attributed to the innate immune system. This study also demonstrated only trace amounts of phage in blood and organs of the mice that received phage by oral administration.

Based on the results of these experiments demonstrating the ability of the mammalian host defense systems to remove phage, a serial-passage selection method was developed to identify phage variants with a capacity to remain for longer periods in the circulatory system (50). This system was used in mice to select two *E. coli* phage λ variants that demonstrated 16,000- and 13,000-fold greater capacities to remain in the circulation by evading the animal's host defense systems; similar results were obtained with the *Salmonella typhimurium* P22 phage. These long-circulating mutant phages were demonstrated to be of value in treating animals with bacterial infections. In these experiments, there was less morbidity in *E. coli*-infected mice treated with the long-circulating λ phage mutants. In similar experiments conducted with *S. typhimurium* P22 phage, mice infected with *Salmonella typhimurium* also displayed less morbidity and mortality when they were treated with the long-circulating mutant P22 (50).

The experiments described above suggested that the loss of phage was due to interactions with the RES. In addition, innate immune system blood factors have also been found to be of importance. Sokoloff et al. (70), using a T7 phage peptide-display library (see chapter 44) found a correlation between the peptides displayed and survival of the phage in the rat circulatory system. Phage displaying C-terminal lysine or arginine residues had longer circulating half-lives. In addition, in rat serum T7 phage inactivation was associated with complement activation. The T7 phage displaying C-terminal lysine or arginine residues were found to be protected from this complement-mediated inactivation by binding to the C-reactive protein which is normally elevated in rats and mice. However, in human serum, phage resistant to inactivation were found to display peptides containing tyrosine residues, not lysine or arginine as in the rat experiments. In human serum the protective protein may be α_2-macroglobulin and not C-reactive protein, as found in the rat and mouse serum experiments, because in contrast to the rat, C-reactive protein is not normally elevated in human serum (70). These T7 phage peptide-library experimental results may also help to explain the finding that there was substitution of a lysine for a glutamic acid residue in the E capsid protein in the long-circulating mutant λ phage used as an antibacterial therapeutic agent in mouse experiments (50).

In addition to selecting phage that can remain longer in the circulatory system, it has been possible, by using phage display libraries, to select phage that display specific peptide sequences that appear to influence the binding or uptake of phage by the vascular endothelium in specific regions of the body (76). This in vivo screening method has also been employed on one patient in an effort to develop a molecular map of the human vasculature (4). The ability to target specific regions of the body may be useful for the treatment of localized infections.

In some infections, the pharmacokinetics of the whole organism may be secondary to the ability to deliver phage intracellularly. This would be particularly important in diseases such as tuberculosis in which the intracellular infection of macrophages can serve as a reservoir for spread of the infection throughout the body. In this regard, Broxmeyer et al. (12) have demonstrated that it is possible to deliver the lytic phage TM4 to intracellular locations in macrophages by using *Mycobacterium smegmatis*, an avirulent mycobacterium, as a vector. In their experiments they showed that such treatment could reduce the titers of Mycobacterium avium or *Mycobacterium tuberculosis* in infected macrophage cultures (12). See chapter 38 for a review of *Mycobacterium* phages.

Another unique feature concerning the pharmacokinetics of phage, unlike most other therapeutic agents, is that phage contain a genome. There is evidence that phage genomes can gain direct entrance to mammalian cells. It has been reported that phage genomic fragments have been found in mammalian cells following oral exposure to phage DNA (21, 65). M13 and λ phage DNA were found, using polymerase chain reaction, in the cells of the gastrointestinal tract, peripheral white bloods, and the cells of the liver and spleen. Phage DNA could be detected for up to 24 hours in the spleen and liver following a single oral dose of phage DNA. However, when phage M13 DNA was fed daily for 1 week, Doerfler et al. (21) were able to isolate clones containing M13DNA from the mouse spleen. One of these clones contained a 1299 nucleotide fragment of M13 DNA covalently linked to an 80 nucleotide DNA segment that had 70% homology to the mouse IgE receptor gene (64). In addition, when pregnant mice were regularly fed phage M13 DNA, evidence of M13 DNA could be detected in the fetuses with in situ hybridization methods. In some rare fetal cells this M13 DNA appeared to be associated with the chromosomes (65). There have also been reports of phage-induced enzyme activity in mammalian cells, albeit at low levels, following exposure to phage or phage DNA (34, 47). The integration of phage DNA into the genomes of mammals, as the result of phage therapy for an infection, might result in the loss of heterozygosity of tumor suppressor genes. However, the effects from such events are probably minimal given that phages are associated with bacteria in our colon, nose, throat, and skin throughout our normal life span. While phage gene delivery and expression in mammalian cells may normally be rare, phage are currently being genetically engineered to enhance these capacities. Such engineered phage may be able to serve as vectors for targeted gene delivery in mammalian cells (40).

Phage in Mammalian Host Infected with Bacteria

For many pharmacological agents, information on drug distribution and clearance would be sufficient. However, phages are not passive pharmacological agents. They are capable instead of exponential growth as is the infectious agent, the bacterium. A full knowledge of the pharmacokinetics of phage antibacterial therapy requires knowledge of three dynamic components and their complex interactions: the infected human, the infecting bacterium, and the phage. Of these three dynamic components two of them, the bacterium and the phage, are capable of exponential growth during the course of the infection and its treatment (see chapter 5 for an overview of the dynamic interactions of two of these components, the phage and bacterium).

One of the first researchers to recognize the need for quantitative data to determine whether phage can sustain exponential growth in vivo, Rene Dubos, made use of an animal infectious diseases model employing intracerebral injections of *Shigella dysenteriae* (22). The data obtained from these experiments demonstrated the multiplication of phage, in infected animals treated with phage, at the site of infection, the brain (figure 48-3) Concurrently he and his coworkers demonstrated that phage treatment was capable of rescuing infected animals (survival of untreated animals was 3.6% while survival of phage-treated animals was 72%). These researchers also showed that heat-inactivated phage provided no protective effects unless the heat-inactivated phage preparations were given days before the bacterial infection. They suggested that this protective effect may have been due to the activation of antibacterial immunity by bacterial products present in the phage lysate.

Thirty-nine years after the publication by Dubos et al. (22), Smith and Huggins reported results in similar experiments (67). In one of their experiments mice were infected intracerebrally with *E. coli* K1 and in another they were infected intramuscularly. In both experiments phage were administered intramuscularly. The results of one of these studies, in which the mice were infected intracerebrally, is illustrated graphically in figure 48-4. As in the Dubos et al. (22) study, the phage levels were highest in the bacteria-infected tissue, the brain. The phage levels fell as the bacterial levels in the brain decreased. Unfortunately, all the data in Smith and Huggins' studies was gathered from only 3 animals at each time point, so no meaningful statistical analysis could be performed. In addition, the graph does not reflect the fact that animals were dying during the course of this experiment. For example, it is not possible by looking at the graph or the data in the tables in the Smith and Huggins (67) paper to recognize that 50% of the untreated animals died by 72 hours and 75% died by 96 hours.

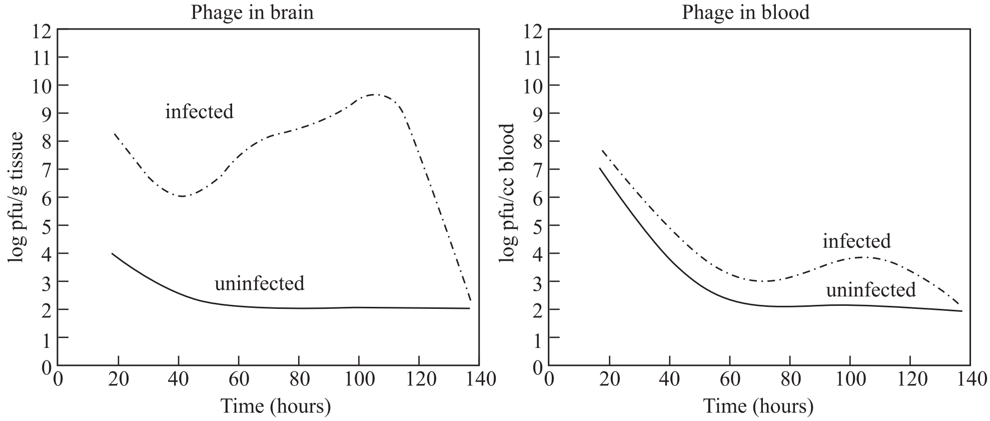


Figure 48-3 Graphical representation of data from the 1943 infectious disease model in which mice inoculated by intracerebral injection of the bacteria *Shigella dysenteriae* (at an LD_{50} level) were compared with uninfected control mice. All the mice in this experiment were injected with 10^9 plague forming units (PFU) of phage intraperitoneally which was administered at the same time as the bacterial inoculation. The bacteriophage level in the blood of the uninfected animals was compatible with the dilution of the phage concentration in the total fluid volume of the mouse and the lower levels in the brain reflect the relatively smaller blood content in the brain. However, in the infected animals the phage particles are observed to increase at the site of the infection, the brain, while the blood levels of phage appear to be a "reflection of the events occurring in the brain."

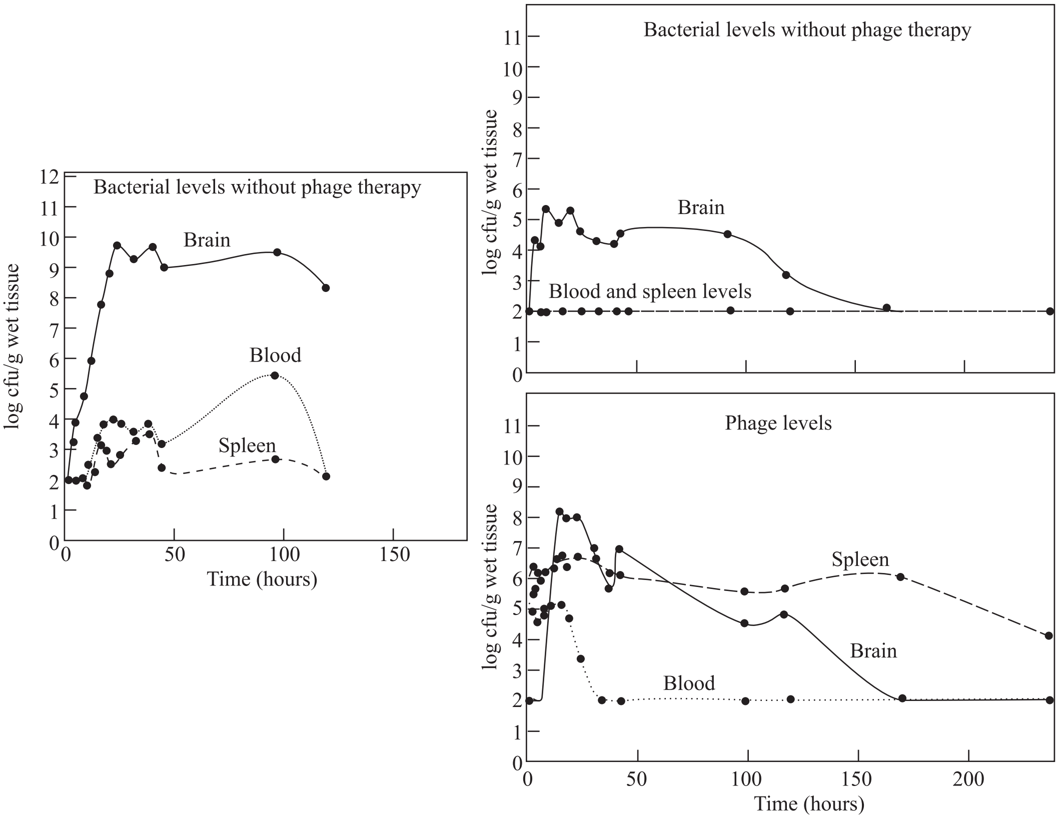


Figure 48-4 Graphical representation of data from phage therapy trial. The data presented are those of Smith and Huggins (67). In this set of experiments all the animals received and intracerebral inoculation of 5×10^2 colony forming units (CFU) of an *E. coli* K1. The animals treated with phage were injected with 3×10^8 plaque forming units (PFU) of phage intramuscularly (into the gastronemius muscle) at the same time as the bacterial inoculation. These graphs were derived from the data published in tables 9 and 10 in Smith and Huggins' paper.

Despite these problems this work has stimulated considerable interest and analysis.

Levin and Bull (43) developed a formal mathematical model based on data from the Smith and Huggins (67) study. Analysis based on this model resulted in their suggestion that the following four elements are critical for successful phage therapy:

1. How "aggressively" phage can lyse bacterial cultures, as determined by adsorption rate, burst size, and latent period is a critical factor. The particular phage that Smith and Huggins found to be most effective therapeutically was both specific against the K1 capsule and known to lyse bacteria more rapidly in culture than a non-K1 specific phage. (See chapter 5 for a general discussion of these various phage growth parameters in terms of phage ecology and the impact of phage on bacteria.)
2. The limited efficacy of antibiotics may be due to the fact that antibiotics decay in the animal whereas the phage (if a sufficient population of bacteria is present) will multiply and increase.
3. Even though phage are capable of multiplication, the initial dose of phage must be sufficient to control the bacterial population before it reaches a lethal threshold.
4. The virulence of the bacterium (how rapidly it can multiply in a host) is also relevant. A more virulent bacterial strain will require either a higher dose of phage or a more virulent phage if the infection is to be countered. The fact that the phage-resistant E. coli that arose following phage therapy were E. coli $K1^-$ and therefore less virulent may or may not have played a role in Smith and Huggins' investigations, but it clearly was not favorable to the bacteria. In contrast, antibiotic-resistant bacterial-mutants are usually *not* less virulent.

The importance of modeling the therapeutic use of phage was also stressed by Payne and Jansen (58). Their model includes terms for the loss of phage due to interaction with mammalian systems, such as the RES of the innate immune system. Studies using their model suggest that the use of antibiotics may at times interfere with phage therapy.

While these models and the data of Smith and Huggins provide some insights into the pharmacokinetics of phage therapy, it is imperative that statistically correct experiments be performed so that more accurate modeling and pharmacological planning can be developed for the therapeutic use of phage. Recently developed methods of visualizing bacterial infections in live animals using bioluminescent strains of bacteria may help in this endeavor (27). In this method bacteria containing a *lux* transposon cassette provide bioluminescent bacteria that can be followed in live mice by the use of a high-sensitivity charge-coupled device camera. The technique has been used to follow pneumococcal infections in the lungs of live mice. By incorporating an appropriate cassette into a phage it should also be possible to follow phage interactions with bacterial infections in vivo (see chapter 46 for review of phage-based reporter systems).

Immunogenicity of Phage

Phage and the Innate Immune System

The immune response is dependent on two components: the innate and the adaptive immune systems (46). The adaptive immune system relies on somatic mutations and clonal expansion of T and B cells in response to an infection. Such clonal expansion requires at least 3–5 days to generate a sufficient number of cells to provide an effective level of antibodies. In contrast, the innate system is dependent on evolution for the development of its functions and it is inherited in a Mendelian fashion. It includes antimicrobial peptides, the alternative complement pathway and phagocytes, including those of the organs of the RES, primarily the liver and spleen. It is the innate immune system that first interacts with a foreign body such as a phage when the animal or person has had no prior exposure to this agent. The same mechanism resulted in the rapid loss of λ phage injected into the circulatory system of germ-free mice in Geier et al.'s (30) experiments, as these mice had no detectable antibody response to phage λ. Likewise, in experiments using T4 phage as a probe of the innate immune system, it was found that the liver phagocytosed more than 99% of that phage within 30 minutes after inoculation and that it removed 12 times as much phage as the spleen, as measured by the uptake of ^{51}Cr-labeled phage (35).

To study the role of blood components of the innate immune system, Sokoloff et al. (70) preinjected rats with GdCl to inhibit phagocytosis by macrophages prior to the administration of T7 phage. In these experiments the PFU of phage in the circulatory system decreased by 95–99% within 5 minutes. As only 10% of the PFUs could be detected in the liver, 1% in the spleen, and less than 1% in the kidneys, lungs, heart, and skeletal muscles it was concluded that most of the phage was inactivated in the blood. This finding was supported by the fact that the half-life of the phage incubated in rat serum at 37°C for 30 minutes was determined to be less than 3 minutes.

Complement was shown to play a major role in this phage inactivation by experiments in which complement activity was inhibited by cobra venom factor (CFV). When CVF was injected intraperitoneally 20 hours before phage injection the loss of phage from the rat circulatory system was significantly reduced. In experiments in vitro the recovery of phage after a 30 minute incubation in rat serum containing CVF was 50%, compared with less than 1% when CVF was not present.

As previously mentioned, by selecting T7 phage from a T7 phage peptide library for its ability to remain for longer periods of time in the rat circulatory system sokoloff et al. found that the long-circulating trait was peptide-specific (70). These long-circulating T7 phage were found display peptides with either a C-terminal lysine or arginine residue. Such peptides appear to protect the phage from complement-mediated inactivation in rat or mouse serum by binding to a serum protein. In Sokoloff et al.'s rats this "protection" protein is was C-reactive protein, which is normally elevated in rats and mice (70). This study is consistent with the prior study by Merril et al. (50) in which a λ phage mutant with a capacity for long circulating times in the mouse was found to have a substitution of a lysine for a glutamic acid in a major capsid protein. It should be noted that Sokoloff et al. found that the protective protein in human serum may be α_2-macroglobulin, rather than C-reactive protein, which has this function in rats and mice. This may explain why phage with tyrosine residues in the displayed peptides were protected in human sera rather than those displaying C-terminal lysine or arginine residues as in the rat sera experiments (70). Experiments such as these suggest that it may be possible to select and/or engineer phage for phage therapy with resistance to inactivation by components in the innate immune system.

Phage and the Adaptive Immune System

Many phage are also potent activators (antigens) of the adaptive immune system. For the past three decades Ochs et al. (56, 57) have made use of this capability of phage ϕX174 to probe the human immune system. In normal individuals injected phage ϕX174 is cleared within 3 days and a primary IgM response can be observed that peaks 2 weeks after immunization. If a second injection is made 6 weeks after the primary immunization, IgM and IgG antibody titers increase and peak within 1 week and subsequent phage injections result in further increases in the IgG titers (57). Patients with severe combined immune deficiency, characterized by absence of both B and T cell functions, display a prolonged clearance of phage, with phage present up to 4–6 weeks after the initial injection. In addition, these severe combined immune deficiency patients do not develop a detectable antibody response to phage. Ochs et al. also noted that while ϕX174 phage is a potent antigen it causes no recognized toxic effects in man (16, 56).

Similar activation of the adaptive immune system was observed in mice inoculated with an antivancomycin resistant enterococcus phage, ENB6 (10). After the third in a series of five monthly injections of phage ENB6, titers of IgG and IgM increased above background 3,800-fold and 5-fold, respectively, and IgG levels did not change substantially after the third injection. No anaphylactic reactions, changes in core body temperature, or other adverse events were observed in the mice over the course of these multiple injections of phage. It may be possible to develop phage that are less antigenic by using phage peptide libraries or affinity matrixes made up of antibodies from human serum. This type of approach has already been initiated to attempt to modulate the immunogenicity of therapeutically important enzymes (36).

Preparation of Phage for Therapeutic Usage

Early phage therapy applications used phages that were purified by filter sterilization. This method has proven to be insufficient because bacterial debris, including bacterial exo- or endotoxins that might be present, can pass through such filters. These contaminants can result in increased morbidity or in some cases mortality. For example, in a recently published study the intraperitoneal inoculation of mice, with 10^9 PFU of filter-sterilized λ phage lysate, grown on *E. coli*, produced a mild reaction (ruffled fur). However, all the mice injected in a similar manner with P22 phage lysate, grown on *Salmonella typhimurium*, died within 12 hours after inoculation. The endotoxin levels in these preparations was 5×10^4 and 5×10^5 endotoxin units (EU)/ml, respectively, as determined by limulus amebocyte lysate assay (50). Additional purification by techniques such as equilibrium density centrifugation with cesium chloride can separate phage particles from debris, including toxins, that does not have the same specific buoyant density as the phage particles. Such centrifugation was able to reduce the endotoxin levels in the phage preparations discussed above to 0.3×10^1 and 1×10^3 EU/ml, respectively. In contrast to the problems noted with the filter-purified phage preparations, no adverse effects were noted in mice inoculated intraperitoneally with phage preparations purified by cesium chloride equilibrium density centrifugation (50). Phage have also been purified by precipitation with ammonium sulfate followed by separation through anion exchange columns (78). Phage prepared in this manner were administered to animals without any noticeable ill effects. In addition, Ochs et al. (56) used this phage purification method in a number of their human protocols.

Testing for adverse effects associated with phage preparations should not be limited to observations of healthy animals. Animals that are stressed may have a lower tolerance to endo- and exotoxins. In a recent study a lower survival rate was observed in bacteremic mice treated with a phage strain (known to have no in vitro activity against the bacteria associated with the bacteremia) (10). In this study, while the highest doses of the phage preparation produced no reported adverse effects in healthy animals an increased mortality was observed in bacteremic stressed mice. This effect was shown to be dependent on the phage dose, suggesting that stressed animals may be more sensitive to the phage itself or to trace amounts of endo- and exotoxins present in the phage preparations than are normal animals.

These experiments suggest that the presence of toxins in early phage preparations may explain some of the catastrophic results reported in early attempts to use phage to treat bacterial infections in humans. In one such example, reported in 1932, a phage strain was found that could lyse broth cultures of plague (*Yersinia pestis*) in less than 2 hours. However, when a filter-sterilized lysate containing this phage stain was injected into rabbits that had been experimentally infected with *Y. pestis* the mortality increased to levels above that found in infected rabbits that were not treated with phage. Furthermore, when this phage preparation was then used to treat 33 human patients they all died (mortality from plague is normally 60–90%) (52).

While the omission of purification processes may result in increased levels of contamination, including toxins, the overzealous addition of agents to assure that there are no active bacteria present in phage preparations can also be detrimental. The association of a "weak" phage preparation and the presence of organomercury compounds was made in a 1932 study of commercial phage preparations from a major US pharmaceutical company (74).

The problems associated with the production of phage for clinical use are not insurmountable, as evidenced by over three decades of phage use in humans (56, 57). In addition, animal experiments have provided evidence that relatively simple phage purification processes, such as cesium chlorided equilibrium density centrifugation, can significantly reduce animal morbidity and mortality.

Methods for the Rapid Determination of Phage Specific for Infecting Pathogen

When a clinician is confronted with a patient with an infectious disease the prudent course of action requires the determination of the identity of the infectious agent. This task can often be time-consuming and laborious, involving isolation and identification of the causative agent. Often, given the time needed to make such a determination, physicians use their best judgment to choose and administer a relatively broad-range antibiotic that is effective for the suspected bacterial strain while they wait for the culture- and antibiotic-sensitivity results. In contrast, if phages are to be used in place of antibiotics it is critical to actually determine the strain of phage to be used, given the generally narrow host range displayed by most phage. Such determinations using current technology could easily take days to perform. If phage are to be used as therapeutic antibacterial agents then a rapid and inexpensive method to determine the nature of the infectious bacteria and its phage susceptibility is needed.

One approach that can be taken to this problem is based on the use of modified phage containing reporter genes (see chapter 46). In this method phage are first isolated and identified as being potential therapeutic agents for a particular species or strain of bacteria. These phage are genetically engineered to encode a reporter gene such that a characteristic color or marker will be produced when the phage infects the specific bacterial strain that is susceptible to that phage. For example, if different strains of phage carrying the luciferase reporter gene were placed in a multiwell plate, along with an aliquot of a clinical isolate, the emission of light from any of the wells would serve to identify the bacterial strain in the clinical isolate as well as the phage strain that may be used to treat the bacterial infection. Such testing could be performed in hours, instead of the days that traditional culturing methods require. This approach has been used with phage carrying the luciferase reporter gene to detect *Listeria* contamination in foods (45). A similar approach has also been developed for a rapid and relatively inexpensive diagnostic test for *Mycobacteria* infections in patients suspected to have tuberculosis (15). Alternatively, given that lysis of bacterial strains by phage will result in the discharge of adenylate kinase which can convert ADP in the reaction mix to ATP, and that luciferin/luciferase can utilize the ATP for light emission, placing luciferin/luciferase in the incubation mixture will serve to identify the organism and the appropriate phage without the need to genetically endow the phage with a luciferase reporter gene (66).

Recent advances in mass spectrometry may also provide fast methods for the identification of bacterial strains. It is now possible to rapidly identify and type bacterial strains based on their lipid, protein, and nucleic acid mass fragment "fingerprints" (41, 79). While mass spectrometry is currently being developed to identify bacterial strains it might also be possible to use this approach to determine whether the bacteria are susceptible to a given phage. However, such information is not currently available and it may be impractical to gain sufficient knowledge of bacterial mass "fingerprints" to accurately determine when a particular phage could be useful. Alternatively, one could use phage gene-product expression for the development of markers both for bacterial identification and as an indicator of phage susceptibility. In this approach one could use mass spectrometry. First a clinical sample would be placed in growth medium to amplify the bacteria, followed by infection with selected "therapeutic" phage strains. If the infecting bacteria were susceptible to the phage, mass spectrometry would then detect signature fragments of proteins that are expressed only when infection of the bacteria occurs by a specific phage. Signature fragments, that are not part of the phage virion, would be generated from phage-specific proteins such as RNA polymerase, regulatory protein, or lytic enzyme. This approach could be used so that the generation and detection of phage-specific products serves the same role as "reporter" gene products in a manner analogous to the detection of marker gene products described above.

DNA microarrays, in conjunction with polymerase chain reaction amplification, are also being developed for the rapid diagnosis of bacterial strains. This technology can be used to determine the susceptibility of bacterial strains to certain antibiotics (31). In principle it may be possible to also develop such a method for the determination of bacterial strains and their phage susceptibility.

Therapeutic Use of Phage Products and Components

Phage gene products and components may also serve as antibacterial therapeutic agents. While such applications lose the exponential growth capacity of phage they may still be highly effective. For example, it has been suggested that phage-encoded polypeptides could be developed into a new class of antibiotics (9). This suggestion is based on the recognition that the small-genome phages ϕX174 and Qβ encode polypeptides that interfere with bacterial wall biosynthesis and that such inhibition can results in bacterial lysis (see chapters 11 and 15, respectively, for reviews of ϕX174 and Qβ biology). In another use of phage gene products it has been demonstrated that phage-encoded endolysins, which disrupt the peptidoglycan matrix of the bacterial cell wall, and phage holins, which permeabilize bacterial membranes, can also serve as effective antibacterial agents (see chapter 10 for review of phage lysis proteins). A single dose of a phage lysin specific for streptococci groups A, C, and E was capable of clearing these bacteria both in vitro and in vivo in mouse upper respiratory infections (54). As this lysin has little if any affect on other commensal organisms, it should be less disruptive of the oral and upper respiratory ecology than most antibiotic treatments. A similar result was obtained when a phage encoded enzyme, PlyG lysine (enclosed by the γ phage of *Bacillus anthracis*), was used to rescue mice infected with *Bacillus cereus*, a bacterial strain that is closely related to *Bacillus anthracis* (66). No resistant *B. cereus* strains were detected following such treatment. In addition, as ATP is released when PlyG lysine destroys *B. anthracis*, this enzyme in conjunction with luciferin/luciferase can also be used to rapidly detect γ-sensitive bacilli and their germinating spores. In this application, spores are detected by first immobilizing them on filter membranes. They are then incubated with germinant and treated with PlyG lysine and luciferin/luciferase. Emitted light is detected using a hand-held luminometer. This system was able to detect as few as 100 spores (66).

Phage-encoded lysin enzymes have also been used prophylactically. Gaeng et al. developed a bacterial strain that secretes the functional bacteriophage lysin enzymes Ply511 and Ply118. They used this bacterial strain to eliminate *Listeria monocytogenes* from dairy starter cultures (used in the production of cheese) (29).

Concluding Remarks

While phage therapy has been employed continually since the initial discoveries of these viruses at the beginning of the twentieth century, these clinical applications have never faced the scrutiny now required in countries that require certification of pharmacological agents. Such certification is based on the results of studies of efficacy and pharmacokinetics in animals and humans. As discussed, there are a number of historical reasons for this deficiency including the overshadowing discovery of the antibiotics with their broader antibacterial host range. However, phage deserve careful review as they may provide ideal therapeutic agents for the treatment of emerging antibiotic-resistant bacterial strains and, as Lederburg (42) suggested, for the treatment of epidemics such as cholera in refugee camps. They may prove especially useful in agricultural applications where their high specificity can be used to treat a bacterial infection without disturbing the larger ecological systems, as is often the problem with broad bacterial host range antibiotics (42). These suggestions are strengthened by the recent observations that many antibiotic-resistant bacterial strains are arising through clonal selection. In recognition of this growing problem, the US Food and Drugs Administration (FDA) recently announced that it is re-evaluating livestock antibiotics currently on the market. In this regard, the FDA is now requiring manufacturers of proposed livestock pharmaceuticals to determine whether newly proposed antibiotics will be associated with the emergence of pathogenic organisms with resistance to drugs currently in use for the treatment of human diseases (37). In a comment in *Nature* concerning presentation at the American Society for microbiology meeting in Salt Lake city in 2002, knight noted the lack of genetic variability in antibiotic-resistant bacteria. He cited evidence presented by Klugman that only 10 strains of pneumococcus are associated with 75% of the cases of antibiotic-resistant childhood pneumonia and one half of these cases are caused by a single strain, "Spain 23-E." Similar results were reported at a meeting by Herminia de Lencastre for a study of methicillin-resistant *Staphylococcus aureus* (MRSA). In this study only Five strains of MRSA were found in 70% of 3000 clinical isolates from 14 countries (38). This lack of genetic variability in antibiotic-resistant bacteria suggests that such pathogenic bacteria may offer ideal targets for phage therapy.

In addition to the potential of phage for the treatment of antibiotic-resistant bacterial infections, phage with their generally narrow host range may be better suited than the currently employed antibiotics for a number of clinical applications. For example, in the treatment of bacterial infections, phage can be targeted toward the specific pathogen without disturbing complex bacterial ecological systems such as those associated with the human gastrointestinal system. Applications of phage to treat infections may eliminate the iatrogenic effects of antibiotics such as the antibiotic-related diarrhea diseases that range from "nuisance" diarrhea to colitis associated with *Clostridium difficile* infections (8).

Despite the clear need for, and in some cases advantages, of, phage therapy there may be some concern over the view of regulatory agencies concerning approval of phage as an antibiotic therapy. However, these agencies are acutely aware of the problems associated with antibiotic-resistant organisms and the need for new approaches to this problem. As for the safety of phage therapy, humans are normally in contact with phage throughout their life, given the complex interactions of bacteria and these viruses in the colon and upper respiratory system, and on the skin. In this regard, many of the phages in current collections were isolated from human waste. While some phage carry harmful genes, it should be possible to eliminate these phage or those genes from our collections of therapeutic phage. In addition, it should be noted that Hans Ochs and his colleagues have been using phage as a means to determine the extent of immune deficiencies and as a probe of the immune system in human studies for the past three decades (56, 57). One additional fact that may be of interest is that many vaccines for human consumption were found, in the 1970s, to be contaminated with phage (from contaminated fetal calf serum used to produce these vaccines). Despite this contamination president Nixon issued an Executive order to permit their continued use (48, 49).

Development of therapeutic phage will require a commitment to fulfill the scientific requirements required of current pharmaceutical agents. In this effort the years of experience gained from the use of phage to discover many of the basic tenets of molecular biology should prove to be an asset. This information — in addition to the encouraging results of recent controlled animal experiments demonstrating the capacity of phage to rescue animals with life-threatening infections — suggests that such an effort may result in the development of needed antibacterial therapeutic agents.

References

1. Alisky, J., K. Iczkowski, A. Rapoport, and N. Troitsky. 1998. Bacteriophages show promise as antimicrobial agents. J. Infect. 16 5–15.
2. Altschul, S. F., G. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403–410.
3. Appelmans, R. 1921. Le bacteriophage dans l'organisme. C. R. Soc. Biol. (Paris) 85:722–724.
4. Arap, W., M. G. Kolonin, M. Trepel, et al. 2002. Steps toward mapping the human vasculature by phage display. Nat. Med. 8:121–127.
5. Asheshov, I. N., J. Wilson, and W. W. C. Topley. 1937. The effect of an anti-vi bacteriophage on typhoid infection in mice. Lancet I:319–320.
6. Atanasiu, C., O. Byron, H. McMiken, S. S. Sturrock, and D. T. F. Dryden. 2001. Characterisation of the structure of ocr, the gene *0.3* protein of bacteriophage T7. Nucl. Acids Res. 29:3059–3068.
7. Barrow, P., M. Lovell, and A. Berchieri. 1998. Use of lytic bacteriophage for control of experimental *E. coli* septicemia and meningitis in chickens and calves. Clin. Diag. Lab. Immunol. 5:294–298.
8. Bartlett, J. G. 2002. Antibiotic-associated diarrhea. N. Engl. J. Med. 346:334–339.
9. Bernhardt, T. G., I.-N. Wang, D. K. Struck, and R. Young. 2001. A protein antibiotic in the phage Qβ virion: diversity in lysis targets. Science 292:2326–2329.
10. Biswas, B., S. Adhya, P. Washart, B. Paul, A. N. Trostel, B. Powell, R. Carlton, and C. R. Merril. 2002. Bacteriophage therapy rescues mice bacteremic from a clinical isolate of vancomycin-resistant *Enterococcus faecium*. Infect. Immun. 70:204–210.
11. Boyd, E. F., B. M. Davis, and B. Hochhut. 2001. Bacteriophage–bacteriophage interactions in the evolution of pathogenic bacteria. Trends Microbiol. 9:137–144.
12. Broxmeyer, L., D. Sosnowska, E. Miltner, O. Chacón, D. Wagner, J. McGarvey, R. G. Barletta, and L. E. Bermudez. 2002. Killing of *Mycobacterium avium* and *Mycobacterium tuberculosis* by a mycobacteriophage delivered by a nonvirulent mycobacterium: a model for phage therapy of intracellular bacterial pathogens. J. Infect. Dis. 186:1155–1160.
13. Brüssow, H., and R. W. Hendrix. 2002. Phage genomics: small is beautiful. Cell 108:13–16.
14. Calalb, G. 1925. Action de la bile sur bacteriophage et importance de cette action. C. R. Soc. Biol. (Paris) 92:1442–1443.
15. Carriere, C., P. F. Riska, O. Zimhoney, J. Kriakov, S. Bardarov, J. Burns, J. Chan, and W. R. Jacobs. 1997. Conditionally replicating luciferase reporter phages: improved sensitivity for rapid detection and assessment of drug susceptibility of *Mycobacterium tuberculosis*. J. Clin. Microbiol. 35:3232–3239.

16. Ching, Y.-C., S. D. Davis, and R. J. Wedgwood. 1966. Antibody studies in hypogammaglobulinemia. J. Clin. Invest. 45:1593–1600.
17. Colvin, M. G. 1932. Behavior of bacteriophage in body fluids and in exudates. J. Infect. Dis. 51:527–541.
18. Danese, P. N., L. A. Pratt, S. L. Dove, and R. Kolter. 2000. The outer membrane protein, antigen 43, mediates cell-to-cell interactions within *Escherichia coli* biofilms. Mol Microbiol.37:42–432.
19. d'Herelle, F. 1926. The Bacteriophage and its Behavior, pp. 490–497, 540–549, 540, 541 Williams and Wilkins, Baltimore, Md.
20. d'Herelle, F., and M. L. Rakieten. 1935. The adaptation of a staphylococcus bacteriophage to an artificially produced anti-bacteriophagic serum. J. Immunol. 28:413–423.
21. Doerfler, W., G. Orend, R. Schubbert, K. Fechteler, H. Heller, P. Wilgenbus, and J. Schroer. 1995. On the insertion of foreign DNA into mammalian genomes: mechanism and consequences. Gene 157:241–245.
22. Dubos, R. J., J. H. Straus, and C. Pierce. 1943. The multiplication of bacteriophage in vivo and its protective effects against experimental infection with *Shigella dysenteria*. J. Exp. Med. 20:161–168.
23. Eaton, M. D., and S. Bayne-Jones. 1934. Bacteriophage therapy. JAMA 103: 1769–1776, 1847–1853, 1934–1939.
24. Evans, A. C. 1933. Inactivation of antistreptococcus bacteriophage by animal fluids. Public Health Rep. 48:411–426.
25. Flaherty, J. E., J. B. Jones, B. K. Harbaugh, G. C. Somodi, and L. E. Jackson. 2000. Control of bacterial spot on tomato in the greenhouse and field with H-mutant bacteriophages. Hortscience 35:882–884.
26. Flaherty, J. E., B. K. Harbaugh, J. B. Jones, G. C. Somodi, and L. E. Jackson. 2001. H-mutant bacteriophages as a potential biocontrol of bacterial blight of geraniums. Hortscience 36:90–100.
27. Francis, K. P., J. Yu. C. Bellinger-Kawahara, D. Joh, M. J. Hawkinson, G. Xiao, T. F. Purchio, M. G. Caparon, M. Lipsitch, and P. R. Contag. 2001. Visualizing Pneumococcal infections in the lungs of live mice using bioluminescent *Streptococcus pneumoniae* transformed with a novel gram-positive lux transposon. Infect. Immun. 69:3350–3358.
28. Gabig, M., A. Herman-Antosiewicz, M. Kwiatkowska, M. Los, M. S. Thomas, and G. Wegrzyn. 2002. The cell surface protein Ag43 facilitates phage infection of *Escherichia coli* in the presence of bile salts and carbohydrates. Microbiology 148:1533–1542.
29. Gaeng, S., S. Scherer, H. Neve, and M. J. Loessner. 2000. Gene cloning and expression and secretion of *Listeria monocytogenes* bacteriophage-lytic enzymes in *Lactococcus lactis*. Appl. Environ. Microbiol. 66:2951–2958.
30. Geier, M. R., M. E. Trigg, and C. R. Merril. 1973. The fate of bacteriophage lambda in non-immune germfree mice. Nature 246:221–223.
31. Hamels, S., J. L. Gala, S. Dufour, P. Vannuffel, N. Zammatteo, and J. Remacle. 2001. Consensus PCR and microarray for diagnosis of the genus *Staphylococcus* species and methicillin resistance. Biotechniques 31:1364–1366, 1368, 1370–1372.
32. Ho, K. 2001. Bacteriophage therapy for bacterial infections. Perspect. Biol. Med. 44:1–16.
33. Hollon, T. 2002. Impossible vaccine tames *Staphylococcus aureus*. Scientist 16:24–28.
34. Horst, J., F. Kluge, K. Beyreuther, and W. Gerok. 1975. Gene transfer to human cells: transducing phage lambda *plac* gene expression in GM1-gangliosidosis fibroblasts. Proc. Natl. Acad. Sci. USA 72:3531–3535.
35. Inchley, C. J. 1969. The activity of mouse Kupffer cells following intravenous injection of T4 bacteriophage. Clin. Exp. Immunol. 5:173–187.
36. Jenne, S., K. Brepoels, D. Collen, and L. Jespers. 1998. High resolution mapping of the B cell epitopes of staphylokinase in humans using negative selection of a phage-displayed antigen library. J. Immunol. 161:3161–3168.
37. Kilman, S. 2002. FDA restricts antibiotic use in livestock to protect people. Wall Street Journal [Sept 12]: D3.
38. Knight, J. 2002. Superbugs reveal chink in armour. Nature 417:477.
39. Larkum, N. W. 1932. Bacteriophage in clinical medicine. J. Lab. Clin. Med. 17:675–680.
40. Larocca, D., M. A. Burg, K. Jensen-Pergakes, E. P. Ravey, A. M. Gonzalez, and A. Baird. 2002. Evolving phage vectors for cell targeted gene delivery. Curr. Pharm. Biotech. 3:45–57.
41. Lay, Jr., J. O. 2001. Maldi-Tof mass spectrometry of bacteria. Mass Spec. Rev. 20:172–194.
42. Lederberg, J. 1996. Smaller fleas ad infinitum: therapeutic bacteriophage redux. Proc. Natl. Acad. Sci. USA 93:3167–3168.
43. Levin, B. R., and J. J. Bull. 1996. Phage therapy revisited: the population biology of a bacterial infection and its treatment with bacteriophage and antibiotics. Am. Nat. 147:881–898.
44. Liu, M., R. Deora, S. R. Doulatov, M. Gingery, F. A. Eiserling, A. Preston, D. J. Maskell, R. W. Simons, P. A. Cotter, J. Parkhill, and J. F. Miller. 2002. Reverse transcriptase-mediated tropism switching in *Bordetella* bacteriophage. Science 295:2091– 2094.
45. Loessner, M. J., C. E. D. Rees, A. B. Steward, and S. Scherer. 1996. Construction of luciferase reporter bacteriophage A511::*luxAB* for rapid and sensitive detection of viable *Listeria* cells. Appl. Environ. Microbiol. 62:1133–1140.
46. Medzhitov, R., and C. Janeway. 2000. Innate immunity. N. Engl. J. Med. 343:338–344.
47. Merril, C. R., M. R. Geier, and J. C. Petricciani. 1971. Bacterial virus gene expression in human cells. Nature 233:398–400.
48. Merril, C. R., T. B. Friedman, A. Attallah, M. R. Geier, K. Krell, and R. Yarkin. 1972. Isolation of bacteriophages from commercial sera. In Vitro 8:91–93.
49. Merril, C. R. 1975. Phage in human vaccines. Science 188:8.
50. Merril, C. R., B. Biswas, R. Carlton, N. C. Jensen, G. J. Creed, S. Zullo, and A. Adhya. 1996. Long-circulating bacteriophage as antibacterial agents. Proc. Natl. Acad. Sci. USA 93:3188–3192.

51. Moffatt, B. A., and F. W. Studier. 1988. Entry of bacteriophage T7 DNA into the cell and escape from host restriction. J. Bacteriol. 170:2095–2105.
52. Naidu, B. P. B., and C. R. Avari. 1932. Bacteriophage in the treatment of plague. Ind. J. Med. Res. 19:737–748.
53. Nakai, T., and S. C. Park. 2002. Bacteriophage therapy for infectious diseases in aquiculture. Res. Microbiol. 153:13–18.
54. Nelson, D., L. Loomis, and V. Fischetti. 2001. Prevention and elimination of upper respiratory colonization of mice by group A streptococci by using a bacteriophage lytic enzyme. Proc. Natl. Acad. Sci. USA 98:4107–4112.
55. Nungester, W. J. and R. M. Watrous. 1934. Accumulation of bacteriophage in spleen and liver following its intravenous inoculation. Proc. Soc. Exp. Biol. Med. 31:901–905.
56. Ochs, H. D., S. D. Davis, and R. J. Wedgwood. 1971. Immunologic responses to bacteriophage ϕX174 in immunodeficiency diseases. J. Clin. Invest. 50:2559–2568.
57. Ochs, H. D., S. Nonoyama, Q. Zhu, M. Farrington, and R. J. Wedgwood. 1993. Regulation of antibody responses: the role of complement and adhesion molecules. Clin. Immunol. Immunopathol. 3:S33–40.
58. Payne, R. J. H., and V. A. A. Jansen. 2001. Understanding bacteriophage therapy in a density-dependent kinetic process. J. Theor. Biol. 208:37–48.
59. Ramesh, V., J. A. Fralick, and R. D. Rolfe. 1999. Prevention of *Clostridium difficile*-induced ileocecitis with bacteriophage. Anaerobe 5:69–78.
60. Randall-Hazelbauer, L. and M. Schwartz. 1973. Isolation of the bacteriophage lambda receptor from *Escherichia coli*. J. Bacteriol. 116:1436–1446.
61. Sandmeier, H. 1994. Acquisition and rearrangement of sequence motifs in the evolution of bacteriophage tail fibres. Mol. Microbiol. 12:343–350.
62. Scholl, D., S. Rogers, S. Adhya, and C. Merril. 2001. Bacteriophage K1-5 encodes two different tail fiber proteins allowing it to infect and replicate on both K1 and K5 strains of *E. coli*. J. Virol. 75:2509–2515.
63. Scholl, D., S. Adhya, and C. R. Merril. 2002. Bacteriophage SP6 is closely related to phages K1-5, K5 and K1E but encodes a tail protein very similar to that of the distantly related P22. J. Bacteriol. 184:2833–2836.
64. Schubbert, R., D. Renz, B. Schmitz, and W. Doerfler. 1997. Foreign (M13) DNA ingested by mice reaches peripheral leukocytes, spleen, and liver via the intestinal wall mucosa and can be covalently linked to mouse DNA. Proc. Natl. Acad. Sci. USA 94:961–966.
65. Schubbert, R., U. Hohlweg, D. Renz, and W. Doerfler. 1998. On the fate of orally ingested foreign DNA in mice: chromosomal association and placental transmission to the fetus. Mol. Gen. Genet. 259:569–576.
66. Schuch, R., D. Nelson, and V. A. Fischetti. 2002. A bacteriolytic agent that detects and kills *Bacillus anthracis.* Nature 418:884–889.
67. Smith, H. W., and M. B. Huggins. 1982. Successful treatment of experimental *E. coli* infections in mice using phage: its general superiority over antibiotics. J. Gen. Microbiol. 128:307–318.
68. Smith, H. W., M. B. Huggins, and K. M. Shaw. 1987. The control of experimental *E. coli* diarrhea in calves by means of bacteriophages. J. Gen. Microbiol. 133:1111–1126.
69. Smith, H. W., M. B. Huggins, and K. M. Shaw. 1987. Factors influencing the survival and multiplication of bacteriophages in calves and their environments. J. Gen. Microbiol. 133:1127–1135.
70. Sokoloff, A. V., I. Bock, G. Zhang, M. G. Sebestyen, and J. A. Wolff. 2000. The interactions of peptides with the innate immune system studied with use of T7 phage peptide display. Mol. Ther. 2:131–139.
71. Soothill, J. S. 1992. Treatment of experimental infections of mice with bacteriophages. J. Med. Microbiol. 37:258–262.
72. Soothill, J. S. 1994. Bacteriophage prevents destruction of skin grafts by *Pseudomonas aeruginosa*. Burns 20:209–211.
73. Stent, G. S. 1963. Molecular Biology of Bacterial Viruses, P. 8 W. H. Freeman San Fransisco.
74. Straub, M. E., and M. Applebaum. 1933. Studies on commercial bacteriophage products. JAMA 100:110–113.
75. Summers, W. C. 2001. Bacteriophage therapy. Annu. Rev. Microbiol. 55:437–451.
76. Trepel, M., W. Arap, and R. Pasqualini. 2002. In vivo phage display and vascular heterogeneity: implications for targeted medicine. Curr. Opin. Chem. Biol. 6:399–404.
77. Twort, F. W. 1915. An investigation on the nature of ultramicroscopic viruses. Lancet II:1241–1243.
78. Uhr, J. W., M. S. Finkelstein, and J. B. Baumann. 1962. Antibody formation. III. The primary and secondary antibody response to bacteriophage ϕX174 in guinea pigs. J. Exp. Med. 115:655–670.
79. Van Baar, B. L. M. 2000. Characterisation of bacteria by matrix-assisted laser desorption/ionisation and electrospray mass spectrometry. FEMS Microbiol. Rev. 24:193–219.
80. Van Helvoort, T. 1992. Bacteriological and physiological research styles in the early controversy on the nature of the bacteriophage phenomenon. Med. History 36:234–270.
81. Wagner, P. L., and M. K. Waldor. 2002. Bacteriophage control of bacterial virulence. Infect. Immun. 70:3985–3993.

Index

Acidianus 507
Adsorption 37
  B. subtilis temperate phages 559
  cyanophages 520
  P1 351
  SPP1 333
  T1 212
  T7 281
  N4 303
Antirepressor 583
Antitermination 83
Archaea 499
Assembly 188
  λ 413
  mycobacteriophages 606
  T1 217
  P2 366
  P22 459
  T7 28
Attachment, T5 268
Attachment sites
  B. subtilis temperate phages 557
  Listeria 595
  P2 375

Bacterial pathogenesis 710
Bacteriophage
  ai 642
  c2 group 573
  filamentous 146
  HK022 92
  L2 644
  L3 640
  L51 638
  Lactobacillus 653
  λ 21, 28, 74, 83, 106, 409
  lipid-containing 171
  Listeria 593
  MAV1 646
  MS2 121
  Mu 469
  mycoplasma 636
  N4 302
  N15 448
  936 species 575
  186 365
  P1 of *E. coli* 350
  P1 of mycoplasma 642
  P2 365
  P4 391
  P22 457
  P335 species 576
  PRD1 161
  Q*β* 120
  single-stranded RNA 175
  SPP1 331
  SpV3 641
  SpV4 639
  Streptomyces 621
  T1 211
  T4 225
  T5 268
  T7 277
  *φ*6 197
  *φ*29 315
  *φ*A1122 546
  *φ*C31 624
  *φ*X174 118, 129
  TLS 211
Bacteriophage-resistant strains 670
Benzer, Seymour 6
Bortadella 714
botulism 710
Brenner, Sydney 6
Bronfenbrenner, Jaques 5

Campbell, Allan 6
Capsid, P22 459
Capsid proteins, mycobacteriophages 606
Caudovirales 8, 9
Chemostat 42
Cholera 716
Cloning vectors
  B. subtilis temperate phages 564
  mycobacteriophage 616
  SPP1 344
Clostridium botulinum 710
Corticoviridae 9, 171
Corynebacterium diphtheriae 711
Crenarchaeota 505
Crick, Francis 6
Cryptic prophages, *Listeria* 596
Cyanobacteria 517
Cystoviridae 9, 197, 203

Death raft 117
Delbrück, Max 4
D'Herelle, Felix 4, 8
Diphtheria 711
DNA entry
  PRD1 166
  T1 212
  T5 268
DNA maturation 52
DNA methyltransferases, *B. subtilis* temperate phages 562
DNA packaging 49, 137
  lactococcal phages 585
  Mu 475
  P1 357
  P4 404
  SPP1 340
  T4 244
  T7 289
  φ29 324
DNA penetration 282
DNA polymerase, φ29 321
DNA repair, T4 244
DNA replication 130, 149, 167
  lactococcal phages 583
  λ 429
  N4 309
  N15 452
  P2 369
  SPP1 335
  T1 215
  T4 229
  T5 273
  T7 287
  φ 29 319
DNA translocation 56
  SPP1 340

Early genes, φYe03–12 548
Early gene expression, T7 284
Early transcription
  Mu 480
  P2 367
Ecology 37
  cyanophages 525
  Lactobacillus phages 659
Ecosystem 43
Ellis, Emory 4
Endolysin 106
Evolution 26, 139, 186
  Archaea 511
  lactococcal phages 585
  λ 419
  Listeria phages 597
  marine phages 539
  mu 487
  mycobacteriophages 613
  P2 383
  SPP1 344
  Streptomyces phages 621
  T1 219
  T4 250
Excision
  Mu 483
  mycobacteriophages 615
Exclusion 290

F plasmid 696
Filamentous phage 146, 636
First-step transfer of DNA 268
Frameshifting 215
  mycobacteriophages 609
Fuselloviridae 9

Genetic switch
  Lactobacillus phages 656
  lactococcal phages 580
Genomics
  Listeria phages 595
  mycobacteriophages 602
  prophage 17
  tailed phages 26
  T1 213

Halobacteria 500
Hershey, Alfred 5
Holin 105
Horizontal gene transfer 527
Host genome degradation, T1 215
Host range
  cyanophages 519
  Listeria phages 593
  T7 281
  Yersinia phages 546
Host recognition 549
Host specificity 729

Immunity
  Mu 474
  N15 450
  P4 398
Immunogenicity 735
Induction 71, 627
  λ 417
  P1 354
  P2 383
Injection, P1 351
Inoviridae 9
Intasome, P2 302
Integrase, P2 378
Integration
  Lactobacillus phages 657
  λ 417
  Listeria 595
  Mu 471
  mycobacteriophages 615
  P4 400
Integration vectors 598
Inversion, Mu 484

Jacob, François 5

Lactococcus lactis 572
Late gene expression, lactococcal phages 585
Late gene transcription
  Mu 481
  P1 357
  P2 373
  P4 401
Late genes, T5 273
Lederberg, Joshua 6
Leviviridae 9

Lipid 171
Lipothrixviridae 9
Luria, Salvadore 4
Lwoff, Andre 5, 8
Lysis 104, 138, 179
  inhibition 112
  Lactobacillus phages 658
  lactococcal phages 586
  λ 415, 431
  Listeria phages 594
  mycobacteriophages 611
  N4 311
  P1 358
  P2 374
  SPP1 342
  T1 218
  *φ*29 324
Lysogenic conversion
  λ 411
  Mu 485
Lysogenization, Mu 474
Lysogeny 66
  B. subtilis temperate phages 559
  cyanophages 521
  Lactobacillus phages 656
  λ 416
  Listeria 595
  marine phages 536
  mycophage L2 645
  mycoplasma phage MAV1 647
  N15 450
  P1 353
  P22 458
Lytic enzymes 598
Lytic infection, marine phages 536
Lytic proteins 563

Marine phages 534
Maturation 52
Membrane 165
Methanogens 502
Microbial diversity 539
Microviridae 9
Middle gene transcription, Mu 480
Middle genes, *φ*Ye03–12 548
Morons 251
  λ 425
  P2 384
Morphogenesis
  Lactobacillus phages 658
  lactococcal phages 585
  N4 310
  P1 357
  SPP1 337
  P4 404
  *φ*29 323
Mycobacteriophages 602
Mycoplasma phages 636, 716
Myoviridae 9, 33

N protein 84
Nuclease, T5 272
Nus proteins 84
NUT site 85

Packaging 49, 137
Packaging mechanism, SPP1 341
Pathogenesis 710
Penetration 129
PET vectors 679
Phage amplification assay 705
Phage display
  filamentous 153, 686
  λ 688
  T4 689
  T7 689
Phage receptor 546
  Listeria phages 593
Phage therapy 725
Photosynthesis 522
Plasmaviridae 9
Plasmid
  addiction 356
  maintenance 337
  partition 355
  P4 391
  prophage 448
Podoviridae 9, 33
Pollution 695
Portal
  mycobacteriophages 609
  P22 463
Pre-early proteins 270
Procapsid 51, 132
  SPP1 338
Prohead 51, 132
  *φ*29 324
Promoter
  λ p_L 676
  λ p_R 677
  Mu 472
  SP6 678
  T3 678
  T5 270
  T7 678, 680
  Yersinia phags 547
Prophage
  defective 71
  establishment 80
  genomics 17
  induction 77
  induction, *B. subtilis* temperate phages 561
  Listeria 595
  maintenance 80
Proteases, mycobacteriohages 609
Protein degradation 240
Protein-primed replication 319
Provirus 511
Pseudolysogeny
  cyanophages 522
  marine phages 536
Pseudomonas 715
PUT sequence 92

Q protein 85
Qut site 85

Receptor 165
  c2 phage group 575
  Lactobacillus phages 656
  protein, T5 269

Recombination
  λ 428
  P2 384
  RNA 203
  T4 229
Regulation, mycobacteriophages 614
Reoviridae 203
Replication origin
  c2 phage group 574
  936 species 576
Replicon, P4 393
Reporter phage 597, 702
Repressor 72
  λ 411
  Mu 482
  P2 379
Resolution, P1 dimers 356
Restriction
  cyanophages 521
  mycoplasma phages 648
  P1 352
  Streptomyces phages 628
  T1 214
  T4 240
RNA packaging 198
RNA polymerase
  N4 304
  RNA-dependent 202
  T7 284
RNA replication 177, 181
Rudiviridae 9

Salmonella 715
Satellite phage 391
Scaffold protein 132
  mycobacteriophages 608
  P22 462
Shigella 715
Siphoviridae 9, 27–32
Site-specific recombination 66
  TP901-1 577
Specialized transduction 68
  B. subtilis temperate phages 561
Species diversity 540
Staphylococcus aureus 711
Streptococci 712
Streptomyces phages 621
Sulfolobus 507
Superinfection exclusion 582

Tail assembly
  SPP1 342
  T4 248
Tail fiber assembly, T4 249
Tails, mycobacteriophages 609
Tailspike, P22 464
Tape measure proteins 610
Taxonomy, *Listeria* phages 593
Tectiviridae 9
Telomere, N15 452
Temperate phages
  B. subtilis 559
  Listeria 595
Terminal redundancy, T7 290
Terminal repeat 547
Terminase, SPP1 340
Therapeutic phage factors 728
Thermococcales 504
Thymidylate synthase 563
Transcription
  c2 phage group 573
  N4 304
  936 species 575
  SPP1 334
  Streptomyces phages 627
  T1 214
  T4 229
  φ29 315
  TP901-1 577
Transcription termination 83
  φ29 318
Transcriptional switch, P2 380
Transducing phages, *Listeria* 596, 598
Transduction
  mycobacteriophages 616
  P1 357
  specialized 68
  SPP1 342
  Yersinia phages 545
Transfection, SPP1 343
Translation 177, 678
  T1 214
Translational control 178–180
Translational repressors 239
Transposition 69
  Mu 474–475
Transposon, mycobacteriophages 617
tRNA genes 605
T7-like phages 546
tuberculosis 617
Twort, Frederick W. 4, 8
Typing phage 597

Vibrio cholerae 716
Virion structure
  N4 302
  T4 235
  T7 278
Virulent phages, *Listeria* 597

Woolman, Elie 5

Yersinia phages 545